

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 5, 2005, 07:48:48 : Search time 183.331 Seconds
(without alignments)
2018.911 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRUTMRSRLVSL.....RASFGSVNPRATPRADYTLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	957	ABP73015	Abp73015 Amino aci
2	5134	100.0	957	ADD22921	Add22921 Acidother
3	5134	100.0	957	ADO52313	Ado52313 Acidother
4	4036	78.6	740	ABP73016	Abp73016 Amino aci
5	4036	78.6	740	ADD22926	Add22926 Acidother
6	4036	78.6	740	ADD22923	Add22923 Acidother
7	3920.5	76.3	733	ADO52315	Ado52315 A. cellu
8	2174	42.3	940	ABB99489	Abb99489 Amino aci
9	1680	32.7	726	ADD22927	Add22927 Aspergill
10	1625.5	31.7	838	ADDA2055	Adda2055 Trichoder
11	1588	30.9	818	ADH51581	Adh51581 Trichoder
12	1125	21.9	812	ADH19118	Adh19118 Geotrichu
13	1117	21.8	789	ADH19120	Adh19120 Geotrichu
14	1114	21.7	826	ADH19124	Adh19124 Geotrichu
15	993	19.3	776	ADR90296	Adr90296 Geotrichu
16	979	19.1	756	ADR90298	Adr90298 Geotrichu
17	979	19.1	757	ADR90302	Adr90302 Geotrichu
18	848	16.5	555	ADD24919	Add24919 Xanthomon
19	823	16.0	1228	ABP73029	Abp73029 Amino aci
20	823	16.0	1228	ADH3636	Adh3636 Acidother
21	702	13.7	762	ABP73022	Abp73022 Amino aci
22	702	13.7	762	ADJ38291	Adj38291 A. cellu
23	468	9.1	88	ABP73018	Abp73018 Amino aci
24	468	9.1	88	ADO52317	Ado52317 A. cellu
25	468	9.1	89	ABP73017	Abp73017 Amino aci

26	468	9.1	89	8	ADO52316	Ado52316 A. cellu
27	462	9.0	88	7	ADD22925	Add22925 Acidother
28	462	9.0	89	7	ADD22924	Add22924 Acidother
29	462	9.0	154	6	ABP73020	Abp73020 Amino aci
30	462	9.0	154	6	ADJ38294	Adj38294 A. cellu
31	459	8.9	150	6	ABP73025	Abp73025 Amino aci
32	459	8.9	150	6	ADH3640	Adh3640 A. cellu
33	431	8.4	1121	8	ADP75895	Adp75895 Acidother
34	429	8.4	153	6	ADP71658	Adp71658 A. cellu
35	429	8.4	153	7	ADP75898	Adp75898 Carbohyd
36	429	8.4	1121	6	ABP71656	Abp71656 A. cellu
37	334	6.5	1751	5	AAV13493	Aav13493 Truncated
38	334	6.5	1751	5	AAE16324	Aae16324 Active ce
39	323	6.3	782	2	AAE16625	Aae16625 Cellulase
40	301.5	5.9	1749	4	AAE72314	Aae72314 Glutamic
41	288	5.6	96	5	ABB78539	Abb78539 Gum arabi
42	276	5.4	1426	2	AAV13492	Aav13492 Truncated
43	276	5.4	1426	5	AAE16323	Aae16323 Active ce
44	274	5.3	406	4	ABG27250	Abg27250 Novel hum
45	262	5.1	2468	6	ABU38411	Abu38411 Protein e

ALIGNMENTS

RESULT 1
ID ABP73015 standard; protein, 957 AA.

XX AC ABP73015;

XX DT 03-JUN-2003 (first entry)

XX DE Amino acid sequence of the avicelase AvIII.

XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;

XX KW detergent; pulp processing; paper processing; feed processing; textile;

XX KW cellulose.

XX OS Acidothermus cellulolyticus.

XX FH Key Location/Qualifiers

XX FT Misc-difference 957 /note= "unspecified residue encoded by N"

XX PN WO2003012090-A2.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023818.

XX PR 28-JUL-2001; 2001WO-US023818.

XX PA (MIDE) MIDWEST RES INST.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX DR WPI; 2003-248177/24.

XX N-PSDB; ABG77632.

XX PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful

XX PS for degradation of cellulose or in generating anti-AvIII antibodies for

XX CC purifying recombinant AvIII polypeptides from genetically engineered

XX CC host cells.

XX CC Claim 2; Page 20; 44pp; English.

XX CC The present sequence represents a thermostable avicelase polypeptide,

XX CC designated AvIII. AvIII is a member of the glycoside hydrolase family

XX CC of enzymes, and is a cellulase. AvIII is useful in the conversion of

XX CC biomass to biofuels and biofuel additives. It may be useful in the

XX CC production of detergents, pulp and paper processing, food and feed

XX CC processing and in textile processes. The thermostable AvIII peptide is

CC useful in the degradation of cellulose, and in generating specific anti-
 CC Aviliii antibodies that are useful in purifying recombinant Aviliii
 CC polypeptides from genetically engineered host cells, in detecting Aviliii
 CC polypeptide expression, as well as a reagent tool for characterizing the
 CC molecular actions of the polypeptide. The Aviliii polynucleotide is useful
 CC as a source of probes or primers in various diagnostic assays

XX Sequence 957 AA;

Query Match 100.0%; Score 5134; DB 6; Length 957;
 Best Local Similarity 100.0%; Pred. No. 4.3e-263;
 Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRSENIRLITMRSRRRLVSLAATASFAVAALGVLPALITASPAHAATTOPTWMSNAIG 60
 DB 1 MDRSENIRLITMRSRRRLVSLAATASFAVAALGVLPALITASPAHAATTOPTWMSNAIG 60
 QY 61 GGGFVDGIVFNEGAPGILVYRTDIGMYRMDANGRWIPLLDWVGNNWNGVNGVSIAD 120
 DB 61 GGGFVDGIVFNEGAPGILVYRTDIGMYRMDANGRWIPLLDWVGNNWNGVNGVSIAD 120
 QY 121 PINTKVMVAAGMTNSWDPNDGAILRSSDQATWQITPPLPKGNNMGRGMBERLAVD 180
 DB 121 PINTKVMVAAGMTNSWDPNDGAILRSSDQATWQITPPLPKGNNMGRGMBERLAVD 180
 QY 121 PINTKVMVAAGMTNSWDPNDGAILRSSDQATWQITPPLPKGNNMGRGMBERLAVD 180
 DB 121 PINTKVMVAAGMTNSWDPNDGAILRSSDQATWQITPPLPKGNNMGRGMBERLAVD 180
 QY 181 PNNNDILYFGAPBGKGLMSTDSGATWSQMTNPDVCTYIANPDTTGQSDIQGVVWA 240
 DB 181 PNNNDILYFGAPBGKGLMSTDSGATWSQMTNPDVCTYIANPDTTGQSDIQGVVWA 240
 QY 181 PNNNDILYFGAPBGKGLMSTDSGATWSQMTNPDVCTYIANPDTTGQSDIQGVVWA 240
 DB 181 PNNNDILYFGAPBGKGLMSTDSGATWSQMTNPDVCTYIANPDTTGQSDIQGVVWA 240
 QY 241 FPKSSSLGQASKTIFVGVADPNNPFWMSRDGATQAVGAPGTFPHKGVDPPVNHV 300
 DB 241 FPKSSSLGQASKTIFVGVADPNNPFWMSRDGATQAVGAPGTFPHKGVDPPVNHV 300
 QY 301 YIATSNVTGPRYDGSQDVWKFVTSQWTRISVPSTDTANDYFGVSGLTIDRQHPNTIM 360
 DB 301 YIATSNVTGPRYDGSQDVWKFVTSQWTRISVPSTDTANDYFGVSGLTIDRQHPNTIM 360
 QY 361 VATQISWMPDITIFRSTDGATWTRIMDWTSPNRSIRLYLDISABEPMLEFGQPNPVP 420
 DB 361 VATQISWMPDITIFRSTDGATWTRIMDWTSPNRSIRLYLDISABEPMLEFGQPNPVP 420
 QY 421 SPKLGMDAMALDPNDRMLYGTATLYATNDLTKWDSGGQIHIAPMKGLLEETVND 480
 DB 421 SPKLGMDAMALDPNDRMLYGTATLYATNDLTKWDSGGQIHIAPMKGLLEETVND 480
 QY 481 LISPPGAPLISALGDIAGFTHADYAVPSTIFTSPEVFTGTSVYALNPSIIVRAGSF 540
 DB 481 LISPPGAPLISALGDIAGFTHADYAVPSTIFTSPEVFTGTSVYALNPSIIVRAGSF 540
 QY 541 DPSSQPNDRHVAFTDQGNKMFQGSSEPGVTTGTVAAASADGSRFVWAPDQGVVYAV 600
 DB 541 DPSSQPNDRHVAFTDQGNKMFQGSSEPGVTTGTVAAASADGSRFVWAPDQGVVYAV 600
 QY 601 GPGNMSAASQGVANNOISDRNPTFYALSNSTFPRSDGVTQPVAAAGLPSSGANG 660
 DB 601 GPGNMSAASQGVANNOISDRNPTFYALSNSTFPRSDGVTQPVAAAGLPSSGANG 660
 QY 601 GGNNSAASQGVANNOISDRNPTFYALSNSTFPRSDGVTQPVAAAGLPSSGANG 660
 DB 601 GGNNSAASQGVANNOISDRNPTFYALSNSTFPRSDGVTQPVAAAGLPSSGANG 660
 QY 661 VWFHAVPGKEGDMILAASGLYHSTNGSSWSAITYGSAVNVYFGSAGSPYPAFVV 720
 DB 661 VWFHAVPGKEGDMILAASGLYHSTNGSSWSAITYGSAVNVYFGSAGSPYPAFVV 720
 QY 661 VWFHAVPGKEGDMILAASGLYHSTNGSSWSAITYGSAVNVYFGSAGSPYPAFVV 720
 DB 661 VWFHAVPGKEGDMILAASGLYHSTNGSSWSAITYGSAVNVYFGSAGSPYPAFVV 720
 QY 721 GTIGVYTGAYRSDDCGTTWVLINDQHOYGNMGOAITGDHANLRRVYIGTNGRGIVYDI 780
 DB 721 GTIGVYTGAYRSDDCGTTWVLINDQHOYGNMGOAITGDHANLRRVYIGTNGRGIVYDI 780
 QY 781 GGAPSSSPSPSVSPSASPSLSBPSPSSSPSPSSSPSSSPSSSPSPSPSPSPSPSPSPSP 840
 DB 781 GGAPSSSPSPSVSPSASPSLSBPSPSSSPSPSSSPSPSSSPSPSPSPSPSPSPSPSPSP 840
 QY 841 SASPSSSPSPSSSPSSSPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 900
 DB 841 SASPSSSPSPSSSPSSSP 900

QY 901 SVDLSVTYVRYWFTRDGSSSTLYVNCDMAIGGNIRASFGSVNPAITPATYLIQ 956
 DB 901 SVDLSVTYVRYWFTRDGSSSTLYVNCDMAIGGNIRASFGSVNPAITPATYLIQ 956

RESULT 2

ID ADD22921 standard; protein; 957 AA.

AC ADD22921;

DT 15-JAN-2004 (first entry)

DE Acidothermus cellulolyticus avicelase Aviliii.

KM enzyme; Aviliii; cellulose reduction; agricultural biomass;

XX municipal solid waste; glycoside hydrolase; avicelase.

OS Acidothermus cellulolyticus.

FH Key Location/Qualifiers

FT Misc-difference 957 /label= Unknown

FT /note= "Encoded by N"

PN US2003108988-A1.

PD 12-JUN-2003.

PF 18-OCT-2002; 2002US-0015400.

PR 28-JUL-2001; 2001US-00917376.

PA (DING/) DING S.

PA (ADNEY/) ADNEY W S.

PA (VINZ/) VINZANT T B.

PA (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel ME;

DR WPI: 2003-810853/76.

DR N-PSDB; ADD22922.

PT New isolated thermal tolerant avicelase polynucleotide useful for
 PT detection of a polynucleotide encoding Aviliii and for reducing cellulose
 PT in a starting material, e.g. municipal solid waste.

PS Claim 16; SEQ ID NO 1; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a
 CC thermostable Aviliii polypeptide. The polynucleotide is useful for
 CC detection of a polynucleotide encoding Aviliii. The polynucleotide is
 CC useful for reducing cellulose in a starting material which involves
 CC administering to the starting material, e.g. agricultural biomass or
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
 CC method further comprises administering a second polypeptide molecule
 CC chosen from the glycoside hydrolase family of proteins. The present
 CC sequence represents the amino acid sequence of Acidothermus
 CC cellulolyticus avicelase Aviliii.

XX Sequence 957 AA;

Query Match 100.0%; Score 5134; DB 7; Length 957;
 Best Local Similarity 100.0%; Pred. No. 4.3e-263;
 Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRSENIRLITMRSRRRLVSLAATASFAVAALGVLPALITASPAHAATTOPTWMSNAIG 60
 DB 1 MDRSENIRLITMRSRRRLVSLAATASFAVAALGVLPALITASPAHAATTOPTWMSNAIG 60
 QY 61 GGGFVDGIVFNEGAPGILVYRTDIGMYRMDANGRWIPLLDWVGNNWNGVNGVSIAD 120
 DB 61 GGGFVDGIVFNEGAPGILVYRTDIGMYRMDANGRWIPLLDWVGNNWNGVNGVSIAD 120

QY	121	PINTKVAAVGMYTNSWDPNDCALIRSSDQGA	TWQITLPLPFLGGMGRGMBERLAND	180
Db	121	PINTKVAAVGMYTNSWDPNDGALIRSSDQGA	TWQITLPLPFLGGMGRGMBERLAND	180
QY	181	PNNDDILYFGASGGLMWSSTSGATWQMTNP	PDVGTIYIANPTDTGYQSDIQGVVWA	240
Db	181	PNNDDILYFGASGGLMWSSTSGATWQMTNP	PDVGTIYIANPTDTGYQSDIQGVVWA	240
QY	241	FDKSSSSIGQASKTIFVGVADPNNPFWMSR	DGATWQAVPGAPGTPIPHKGVDPVNNHL	3000
Db	241	FDKSSSSIGQASKTIFVGVADPNNPFWMSR	DGATWQAVPGAPGTPIPHKGVDPVNNHL	3000
QY	301	YIATISNTGCPYDGSSGDVWKFSTVSGTWR	ISFVPESTDPANDYFGVSGLTIDRQHNTIM	3600
Db	301	YIATISNTGCPYDGSSGDVWKFSTVSGTWR	ISFVPESTDPANDYFGVSGLTIDRQHNTIM	3600
QY	361	VATQISWMPDITIIPSTDGGAITWTRIMDWT	SYNRSILRYLDISAEPMILTFGQPNPVP	4200
Db	361	VATQISWMPDITIIPSTDGGAITWTRIMDWT	SYNRSILRYLDISAEPMILTFGQPNPVP	4200
QY	421	SPKLGMDMEAMADIPENSIDRMUYGTGATL	YATNDLTRKMSGQGIHAPVYKGLBETA	VND 4800
Db	421	SPKLGMDMEAMADIPENSIDRMUYGTGATL	YATNDLTRKMSGQGIHAPVYKGLBETA	VND 4800
QY	481	LISPSGAPLISALGDLGFTHADVTAVPST	ITFSTVPTTGSVDYAEALNPSIIVRAGSF	5400
Db	481	LISPSGAPLISALGDLGFTHADVTAVPST	ITFSTVPTTGSVDYAEALNPSIIVRAGSF	5400
QY	541	DPSSQPNDRHVAFSTTGGKMWFGQSRBPG	VTTGGTGAASADGSRFWMAPDPCQPVYAV	6000
Db	541	DPSSQPNDRHVAFSTTGGKMWFGQSRBPG	VTTGGTGAASADGSRFWMAPDPCQPVYAV	6000
QY	601	GFNGNMAASQGVPAANAQIRSDRNPXPT	FYALSNGTFRYSTDGVTQOPVAAGLPSSGAVG	6600
Db	601	GFNGNMAASQGVPAANAQIRSDRNPXPT	FYALSNGTFRYSTDGVTQOPVAAGLPSSGAVG	6600
QY	661	VMFHAVPGKEGDLMLAASGGLYHSTING	SSWSAITGVSSAVNNGFGKSAFGSSTPAV	FVY 7200
Db	661	VMFHAVPGKEGDLMLAASGGLYHSTING	SSWSAITGVSSAVNNGFGKSAFGSSTPAV	FVY 7200
QY	721	GTIGGVTGAYRSPDCCTTWWLINDDOH	QVNMCOALITGDHANLRVYIGNNGRIYV	GDI 7800
Db	721	GTIGGVTGAYRSPDCCTTWWLINDDOH	QVNMCOALITGDHANLRVYIGNNGRIYV	GDI 7800
QY	781	GGAPSGSPSPSVSPASPLSPSPSSSP	SPSPSPSSSPSPSPSPSPSPSPSPSP	8400
Db	781	GGAPSGSPSPSVSPASPLSPSPSSSP	SPSPSPSPSPSPSPSPSPSPSPSPSP	8400
QY	841	SASPSPSSSPSPSSSPSPSPSSSPV	SGVYQYKNNDSAPGDNQIRPGLQVNTGS	9000
Db	841	SASPSPSSSPSPSSSPSPSPSSSPV	SGVYQYKNNDSAPGDNQIRPGLQVNTGS	9000
QY	901	SSVDLSTVYRVWFTDRDGSSTLVYNC	DMAAIGCINIRASFGSVNPTPTADYTLQ	956
Db	901	SSVDLSTVYRVWFTDRDGSSTLVYNC	DMAAIGCINIRASFGSVNPTPTADYTLQ	956
RESULT 3				
AD052313				
ID AD052313 standard; protein; 957 AA.				
AC AD052313;				
DT 15-JUL-2004 (first entry)				
DE Acidothermus cellulolyticus avicelae III (AviIII).				
KM Thermococcal cellulase; avicelase III; AviIII; cellulose reduction;				
KM biomass degradation; ethanol formation; industrial chemical;				
KM fabric treatment; enzyme.				
OS Acidothermus cellulolyticus.				

XX	Key	Location/Qualifiers
EH	Peptide	1..34
FT	Protein	/label= Signal peptide
FT	Domain	35..957
FT	Domain	/note= "Acidothermus cellulolyticus mature avicelase III"
FT	Domain	35..187
FT	Domain	/note = Carbohydrate binding domain type III
FT	Misc-difference	231..870
FT	Misc-difference	/note = GH74 catalytic domain
FT	Misc-difference	957
XX		/label= Unknown
XX		US2004038334-A1.
XX		26-FEB-2004.
PD		28-JUL-2001; 2001US-00917376.
XX		28-JUL-2001; 2001US-00917376.
PR		(DING/) DING S.
PA		(ADNE/) ADNEY W S.
PA		(VINZ/) VINZANT T B.
PA		(HIMW/) HIMMEL M E.
PI		Ding S, Adney WS, Vinzant TB, Himmel ME;
DR		WPI; 2004-203224/19.
XX		N-PSDB; ADO52314.
PT		Novel thermostable AvIII polypeptide of glycoside hydrolase family and isolated from Acidothermus cellulolyticus, useful for degrading treated biomass into simpler forms of carbohydrate.
PS		Claim 12; SEQ ID NO 1; 19pp; English.
XX		The invention relates to a thermostable cellulase enzyme, avicelase III (AvIII) and its nucleic acid sequence. AvIII is useful for reducing cellulose in a starting material. A thermostable AvIII peptide is useful for degrading treated biomass into simpler forms of carbohydrate, which is used in the formation of ethanol or other industrial chemicals. It is also useful for treating fabrics to remove cellulose-containing stains. The present sequence is Acidothermus cellulolyticus AvIII protein.
XX		Sequence 957 AA;
XX		Query Match 100.0%; Score 5134; DB 8; Length 957;
XX		Best Local Similarity 100.0%; Pred No. 4.3e-263; Indels 0; Gaps 0
XX		Matches 956; Conservative 0; Mismatches 0;
QY	1	MDRSENIRLTMSRRLVSLAATASFAVAALGVLPDIAITASPAHAATQPTWNSVAIG 60
DB	1	MRSSEIRLTRTMSRRRLVSLAATASFAVAALGVLPDIAITASPAHAATQPTWNSVAIG 60
QY	61	GGGFVDGIYFNEGAPGILYKRTDYGKTRWDANGRWIPLLDVWGNNMGYNQVSIAD 120
DB	61	GGGFVDGIYFNEGAPGILYKRTDYGKTRWDANGRWIPLLDVWGNNMGYNQVSIAD 120
QY	121	PIINTNVMAAVGMYTNSWDPNDGAILRSSQGAQTWOITPLPFLGGMPPRGKGEILAYD 180
DB	121	PIINTNVMAAVGMYTNSWDPNDGAILRSSQGAQTWOITPLPFLGGMPPRGKGEILAYD 180
QY	181	PNNDNILYFAGBSGKGLMRSTDSGATWSQWNTNPDVGYIYIANPDTTGYQSDIQGVVWA 240
DB	181	PNNDNILYFAGBSGKGLMRSTDSGATWSQWNTNPDVGYIYIANPDTTGYQSDIQGVVWA 240
QY	241	FDKSSSLGQASKTTFIVGVAADPNPNPVFWSFDGATYQAVDGAFTGFIPIHKGVDPPVNHYL 300
DB	241	FDKSSSLGQASKTTFIVGVAADPNPNPVFWSFDGATYQAVDGAFTGFIPIHKGVDPPVNHYL 300
QY	301	YIATSNITGPIYGGSGDWKFSSTISGWTIRISVPSSTDPDANDYFGYSGLTIIDQHPNTIM 360
DB	301	YIATSNITGPIYGGSGDWKFSSTISGWTIRISVPSSTDPDANDYFGYSGLTIIDQHPNTIM 360

Db	301	IIAHSNTGPGYHSGSSGDVWKEFSVTSCSTWTRISRPVSTPDTRANDYFGVSGLTIDRQHNTIM	360
Qy	361	VATQISWMPDITIIIFRSTDGAGATWTRIMDWTSPNRSGLRYVLDISAEPMLTFGVQPNPVP	420
Db	361	VATQISWMPDITIIIFRSTDGATWTRIMDWTSPNRSGLRYVLDISAEPMLTFGVQPNPVP	420
Qy	421	SPKLGMDNBAMAIIDPPNSDRMIYGTGATLYATNDLTKMPSGCGIHIAPMVKGLEETAVND	480
Db	421	SPKLGMDNBAMAIIDPPNSDRMLYGTGATLYATNDLTKMPSGCGIHIAPMVKGLEETAVND	480
Qy	481	LISPSGAPLISALDGLGFTADVAVAVNSTIFTSVPVFTTGTVDAELNPSITIVRAGSF	540
Db	481	LISPSGAPLISALDGLGFTADVAVNSTIFTSVPVFTTGTVDAELNPSITIVRAGSF	540
Qy	541	DPSGQPNDRHVAFSITDGGKNMFQSGEPGGVTTGGTIVASADGSRFYWAPGDPQPVYAV	600
Db	541	DPSGQPNDRHVAFSITDGGKNMFQSGEPGGVTTGGTIVASADGSRFYWAPGDPQPVYAV	600
Qy	601	GFGNSMAASQGVPAANAQITSDRVNPTFYALSNGTYYRSTIDGVTTPQVAAGLPSSGAVG	660
Db	601	GFGNSMAASQGVPAANAQITSDRVNPTFYALSNGTYYRSTIDGVTTPQVAAGLPSSGAVG	660
Qy	661	VMFHAVPGKEBGLMLTAASSGLYHSTNTGGSSMSAITGVSSAVVNYGPGKSAPGSSYPVAVFY	720
Db	661	VMFHAVPGKEBGLMLTAASSGLYHSTNTGGSSMSAITGVSSAVVNYGPGKSAPGSSYPVAVFY	720
Qy	721	GTIGGVTGAYRSDDCGTTWVLINDDQHQYGNMGOALITGDHANLRVYITGNRGCIYVGD	780
Db	721	GTIGGVTGAYRSDDCGTTWVLINDDQHQYGNMGOALITGDHANLRVYITGNRGCIYVGD	780
Qy	781	GCAPSGSPSPSVSPASPSLSP	840
Db	781	GCAPSGSPSPSVSPASPSLSP	840
Qy	841	SASPSPPSSPSPSSPSSP	900
Db	841	SASPSPPSSPSPSSPSSP	900
Qy	901	SSVDLSTVTRVWFPTRDGSSSTLVNCDMAAICGNIRASFGSVNPTPTADTYLQ	956
Db	901	SSVDLSTVTRVWFPTRDGSSSTLVNCDMAAICGNIRASFGSVNPTPTADTYLQ	956
RESULT 4			
ABP73016			
ID ABP73016 standard; peptide; 740 AA.			
XX	XX	ABP73016;	
AC	XX		
XX	XX		
DT	03-JUN-2003	(first entry)	
XX	XX		
DE	Amino acid sequence of the avicelase AvIII catalytic domain.		
XX	KM	Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; bifuel;	
KW	detergent; pulp processing; paper processing; feed processing; textile;		
XX	OS	cellulose.	
XX	OS	Acidothermus cellulolyticus.	
XX	XX		
FN	WO2003012090-A2.		
PD	13-FEB-2003.		
XX	XX		
PF	28-JUL-2001; 2001WO-US023818.		
XX	XX		
PR	28-JUL-2001; 2001WO-US023818.		
XX	XX		
PA	(MIDE) MIDWEST RES INST.		
PI	Ding S, Adney WS, Vinzant TB, Himmel ME;		
XX	XX		
DR	WPI; 2003-248177/24.		
XX	XX		

PT	New thermostable AvIII peptide or from Acidothermus cellulolyticus, useful for degradation of cellulose or in generating anti-AvIII antibodies for purifying recombinant AvIII polypeptides from genetically engineered host cells.
PT	
XX	Claim 6; Page 8; 44pp; English.
PS	
CC	The present sequence is derived from a thermostable avicelase, designated
CC	AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,
CC	and is a cellulase. AvIII is useful in the conversion of biomass to
CC	biofuels and biotfuel additives. It may be useful in the production of
CC	detergents, pulp and paper processing, food and feed processing and in
CC	textile processes. The thermostable AvIII peptide is useful in the
CC	degradation of cellulose, and in generating specific anti-AvIII
CC	antibodies that are useful in purifying recombinant AvIII polypeptides
CC	from genetically engineered host cells, in detecting AvIII polypeptide
CC	expression, as well as a reagent tool for characterizing the molecular
CC	actions of the polypeptide. The AvIII polynucleotide is useful as a
CC	source of probes or primers in various diagnostic assays
XX	
SQ	Sequence 740 AA;
Query Match	78.6%; Score 4036; DB 6; Length 740;
Best Local Similarity	100.0%; Pred. No. 3,6e-205; Mismatches 0; Indels 0; Gaps 0;
Matches 740;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	47 ATTQPYTSMNAVIGGGGFGVDIVFNEGAPGILYVRDITCGMYRMDANGRWIPLLDWVGW 106
DB	1 ATTQGYTSMNVAIGGGGFVDIVFNEGAPGILYVRDITCGMYRMDANCRWIPLLDWVGW 60
QY	107 NNMGINGVSIADPINTKMAAVAGMTTNSMHPNDGAILRSSDGAATWQITPLPFKLAG 166
DB	61 NNMGNNGVSSIADPINTKMAAAGMTNSMPNDGAILRSSDGAATWQITPLPFKLAG 120
QY	167 NNPGRGMGERLAVDDNNNTNIIYFGAPSGGGLRSSTDGATGWSQMTPDVGTIYIANPTDT 226
DB	121 NNPGRGMGERLAVDDNNNDILIYFGAPSGGGLKRSTDSGATGWSQMTPDVGTIYIANPTDT 180
QY	227 TGYOSDIGVWWAFADKSSSLSGASAKTIYGVADPNPVFMFSRDGATWQAVPGAPTGF 286
DB	181 TGYSODICGVWWAFADKSSSLSGASAKTIYGVADPNPVFMFSRDGATWQAVPGAPTGF 240
QY	287 IPHKGVPDPVNHVLIATISNTGGPYDGSSGDVWKFSVTISGWTRISLPVPSDTDYANDYRGY 346
DB	241 IPHKGVPDPNVNHVIATISNTGPSYDGSSGDVWKFSVTISGWTRISLPVPSDTDYANDYRGY 300
QY	347 SGLTIDTROHPNTIMATOISWMPDITIIRSTDGATWTIRIMOWTSYPNRSLRYVLDISAE 406
DB	301 SGLTIDTROHPNTIMATOISWMPDITIIRSTDGATWTIRIMOWTSYPNRSLRYVLDISAE 360
QY	407 PMLTGVGOENPPVPSPFKLGMDDEAAAIPEFNSDRMLYGATGLVLTNDLTTKMSGGQIH 466
DB	361 PMLTGVGOENPPVPSPFKLGMDDEAAALDPFNSDRMLYTGATGLVLTNDLTTKMSGGQIH 420
QY	467 APMVXGLEETA VNDLISPSPGAPLISALGDLGFTHADYTA VPSTIFTPVFETGTSVDY 526
DB	421 APMVXGLEETA VNDLISPSPGAPLISALGDLGFTHADYTA VPSTIFTPVFETGTSVDY 480
QY	527 AELANSIITYRASFPSPSQPNDRHYAFSTDGKRMFGQSEPCGVTTGGTVAA SADGSRV 586
DB	481 AELANSIITYRASFPSPSQPNDRHYAFSTDGKRMFGQSEPCGVTTGGTVAA SADGSRV 540
QY	587 WAPGPBGQVVYA VFGNSMAASOGVPANAQJRSRPVNRKTFYALSNGFYASTDGGVTF 646
DB	541 WAPGPBGQVVYA VFGNSMAASOGVPANAQJRSRPVNRKTFYALSNGFYASTDGGVTF 600
QY	647 QEVAAGLPSSGAVGVWFHAVPKEGDLMLAASSGLYHSITNGSSWSAITGVSAVANVGF 706
DB	601 QVVAAGLPSSGAVGVWFHAVPKEGDLMLAASSGLYHSITNGSSWSAITGVSAVANVGF 660
QY	707 KSAPGSSYPAVNVGTIGCTAYCASDDCGTTWVLINDOHOYGMWGAIITGDHANLRVY 766
DB	661 KSAPGSSYPAVNVGTIGCTAYCASDDCGTTWVLINDOHOYGMWGAIITGDHANLRVY 720

QY 767 YIGTNGRGIVGDIIGAPSG 786
 Db 721 YIGTNGRGIVGDIIGAPSG 740

RESULT 5
 ADD22926
 ID ADD22926 standard; protein; 740 AA.
 AC ADD22926;
 XX
 XX 15-JAN-2004 (first entry)
 DT
 XX Acidothermus cellulolyticus avicelase AvIII catalytic domain.
 DE
 XX enzyme; AvIII; cellulose reduction; agricultural biomass;
 KM municipal solid waste; glycoside hydrolase; avicelase.
 XX
 XX Acidothermus cellulolyticus.
 OS
 XX US2003108988-A1.
 XX
 XX PD 12-JUN-2003.
 XX
 XX PF 18-OCT-2002; 2002US-00155400.
 XX
 XX PR 28-JUL-2001; 2001US-00917376.
 XX
 XX PA (DING/) DING S.
 PA (ADNEY/) ADNEY W S.
 PA (VINZ/) VINZANT T B.
 PA (HIMM/) HIMMEL M E.
 XX
 XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
 DR WPI; 2003-810853/76.
 XX
 XX PT New isolated thermal tolerant avicelase polynucleotide useful for
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose
 PT in a starting material, e.g. municipal solid waste.
 XX
 XX PS Example 2; SEQ ID NO 6; 29pp; English.
 XX
 XX CC The invention relates to an isolated polynucleotide molecule encoding a
 CC thermostable AvIII polypeptide. The polynucleotide is useful for
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is
 CC useful for reducing cellulose in a starting material which involves
 CC administering to the starting material, e.g. agricultural biomass or
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
 CC method further comprises administering a second polypeptide molecule
 CC chosen from the glycoside hydrolase family of proteins. The present
 CC sequence represents the amino acid sequence of Acidothermus
 CC cellulolyticus avicelase AvIII catalytic domain.
 CC
 CC SQ Sequence 740 AA;

Query Match 78.6%; Score 4036; DB 7; Length 740;

Best Local Similarity 100.0%; Pred. No. 3.6e-205; Indels 0; Gaps 0;
 Matches 740; Conservative 0; Mismatches 0;

QY 47 ATTQPTWSNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRMDAANGRIPLLDWGW 106
 Db 1 ATTQPTWSNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRMDAANGRIPLLDWGW 60
 QY 107 NNNGYNGVSIADPINTNKVMAAVGYTNSWDPNDGAILRSSDQATWOTPLPFKGG 166
 Db 61 NNNGYNGVSIADPINTNKVMAAVGYTNSWDPNDGAILRSSDQATWOTPLPFKGG 120
 QY 167 NMGRGGRGERLAVDPNNNDILYFGAPSGKLMRSTDSGATWSOMTNPFDVGTYIANPTDT 226
 Db 121 NMGRGGRGERLAVDPNNNDILYFGAPSGKLMRSTDSGATWSOMTNPFDVGTYIANPTDT 180

QY 227 TGYQSDIQGVVWMAFDKSSSSSLGQASKTIFVGVADPNNPVFWSRSDGATWQAVGAPTYGF 286
 Db 181 TGYQSDIQGVVWMAFDKSSSSSLGQASKTIFVGVADPNNPVFWSRSDGATWQAVGAPTYGF 240
 QY 287 IPRKGVDPVNNHLYIATSNVTGCPYDSSSGDVMKEFSTSGTWTRISLPSSTDTANDYGY 346
 Db 241 IPRKGVDPVNNHLYIATSNVTGCPYDSSSGDVMKEFSTSGTWTRISLPSSTDTANDYGY 300
 QY 347 SGLTIDROHPNTIWAQISWMPDTIIFRSTDGATWTRIMWMTSYPRRSIRVYLIDISAE 406
 Db 301 SGLTIDROHPNTIWAQISWMPDTIIFRSTDGATWTRIMWMTSYPRRSIRVYLIDISAE 360
 QY 407 PMLTFGVQPNPVPBPKLGWMDZAMAIDPENSDBMLYGTGATLYATNDLYTWSGGQIHI 466
 Db 361 PMLTFGVQPNPVPBPKLGWMDZAMAIDPENSDBMLYGTGATLYATNDLYTWSGGQIHI 420
 QY 467 APMTKGLBEETAVNDLISPPSGAPLISALGDIAGTTHADVAVPSTIFTSPPVFTGTSDY 526
 Db 421 APMTKGLBEETAVNDLISPPSGAPLISALGDIAGTTHADVAVPSTIFTSPPVFTGTSDY 480
 QY 527 AELNPSIIIVRGSFDPSSQPNDRHVAFSTDGKMFQGSBFGVTTGTVAAASADGSRFV 586
 Db 481 AELNPSIIIVRGSFDPSSQPNDRHVAFSTDGKMFQGSBFGVTTGTVAAASADGSRFV 540
 QY 587 WAPGDPQPVVYAVGFNSMAASQGVPMNAQIRSDRVNPKTFYALSNCTFRSTDGVTYF 646
 Db 541 WAPGDPQPVVYAVGFNSMAASQGVPMNAQIRSDRVNPKTFYALSNCTFRSTDGVTYF 600
 QY 647 QPVAAGLPSSGAVGMFHAVPKKGDLMLAASGLYHSTNGSSWSALTYGSAVNVGFG 706
 Db 601 QPVAAGLPSSGAVGMFHAVPKKGDLMLAASGLYHSTNGSSWSALTYGSAVNVGFG 660
 QY 707 KSAPGSSYPAFVFWGTIGCVTVGAYRSDDCGTWVLINDOHOYGMWGAITGDHANLRV 766
 Db 661 KSAPGSSYPAFVFWGTIGCVTVGAYRSDDCGTWVLINDOHOYGMWGAITGDHANLRV 720
 QY 767 YIGTNGRGIVGDIIGAPSG 786
 Db 721 YIGTNGRGIVGDIIGAPSG 740

RESULT 6

ADD22923
 ID ADD22923 standard; protein; 740 AA.

AC ADD22923;

DT 15-JAN-2004 (first entry)

DE Acidothermus cellulolyticus avicelase AvIII catalytic domain.

XX enzyme; AvIII; cellulose reduction; agricultural biomass;
 KM municipal solid waste; glycoside hydrolase; avicelase.
 XX
 XX OS Acidothermus cellulolyticus.
 XX
 XX PN US2003108988-A1.
 XX
 XX PD 12-JUN-2003.
 XX
 XX PF 18-OCT-2002; 2002US-00155400.
 XX
 XX PR 28-JUL-2001; 2001US-00917376.
 XX
 XX PA (DING/) DING S.
 PA (ADNEY/) ADNEY W S.
 PA (VINZ/) VINZANT T B.
 PA (HIMM/) HIMMEL M E.
 XX
 XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
 DR WPI; 2003-810853/76.
 XX

PT New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding Aviii and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.

PS Claim 16; SEQ ID NO 3; 29pp; English.

CC The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable Aviii polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding Aviii. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents the amino acid sequence of Acidothermus
CC cellulolyticus avicelase Aviii catalytic domain.

XX Sequence 740 AA;

Query Match 78.6%; Score 4036; DB 7; Length 740;
Best Local Similarity 100.0%; Pred. No. 3.6e-205;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 47 ATTGPTWNSNVAIGGGGFVDGIVNEGAPGILYVRTDIGMYRMDANGRWIPLLDWGM 106
DB 1 ATTGPTWNSNVAIGGGGFVDGIVNEGAPGILYVRTDIGMYRMDANGRWIPLLDWGM 60
QY 107 NNMGYNGVSIADPINTNKVMAAVGMVYTNMWDNDGAILRSSDQATWQITPLPFLKGG 166
DB 61 NNMGYNGVSIADPINTNKVMAAVGMVYTNMWDNDGAILRSSDQATWQITPLPFLKGG 120
QY 167 NMPGRGGERLAVDPNNNDILYFGAPSGKLMRSTSGATWSQMTNFPDVGYIYANPTDT 226
DB 121 NMPGRGGERLAVDPNNNDILYFGAPSGKLMRSTSGATWSQMTNFPDVGYIYANPTDT 180
QY 227 TGYOSDIGVWVWAFDKSSSSSLGQASKTIFVGVADPNNPFWRSRDGATQAVPGATGF 286
DB 181 TGYOSDIGVWVWAFDKSSSSSLGQASKTIFVGVADPNNPFWRSRDGATQAVPGATGF 240
QY 287 IPHKGVDPVNHVLYATSNVTGPGYDSSGDVWKFVSTGSTRISFPVSTDANDYFGY 346
DB 241 IPHKGVDPVNHVLYATSNVTGPGYDSSGDVWKFVSTGSTRISFPVSTDANDYFGY 300
QY 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMWTSYPNRSLRYVLDISAE 406
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMWTSYPNRSLRYVLDISAE 360
QY 407 PMLTFGVQNPVPSPKLGWMDAMADPNSDRMLYGTGATLYATNDLTKMDSGGQIHI 466
DB 361 PMLTFGVQNPVPSPKLGWMDAMADPNSDRMLYGTGATLYATNDLTKMDSGGQIHI 420
QY 467 AMVVKGLEETAVNDLISPPSGAPLISALGPGFTHADTVAVPSTIPTSVEFTTGTSDY 526
DB 421 AMVVKGLEETAVNDLISPPSGAPLISALGPGFTHADTVAVPSTIPTSVEFTTGTSDY 480
QY 527 AELNPSIIVRAGSFDESSQPNDRHVAFTDGGKMFQGSBPGVTTGGTVAASADGSRFV 586
DB 481 AELNPSIIVRAGSFDESSQPNDRHVAFTDGGKMFQGSBPGVTTGGTVAASADGSRFV 540
QY 587 NARPGDGGPVYVAVGFGNSWMAAQGVANAOIRSDRNPRTFYALSNGTYRSTDGVTF 646
DB 541 NARPGDGGPVYVAVGFGNSWMAAQGVANAOIRSDRNPRTFYALSNGTYRSTDGVTF 600
QY 647 QPVAAGLPSSGAVGMFHAVPGKEGDLMLAASGLYHSTGSSWSAITGVSSAVNYGFG 706
DB 601 QPVAAGLPSSGAVGMFHAVPGKEGDLMLAASGLYHSTGSSWSAITGVSSAVNYGFG 660
QY 707 KSAFGSSYPAVFVVTIGVYGA YRSDDCGTTWTLINDDOHQYGNMCOAITGDHANLRV 766
DB 661 KSAFGSSYPAVFVVTIGVYGA YRSDDCGTTWTLINDDOHQYGNMCOAITGDHANLRV 720
QY 767 YIGTNGRGIVYGDIGGAPSG 786
DB 767 YIGTNGRGIVYGDIGGAPSG 786
```

DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 7
AD052315
ID AD052315 standard; protein; 733 AA.

XX AC AD052315;

DT 15-JUL-2004 (first entry)

DE A. cellulolyticus avicelase III (Aviii) GHCA catalytic domain (CD).

KM Thermostable cellulase; avicelase III; Aviii; cellulose reduction;

KW biomass degradation; ethanol formation; industrial chemical;

KM fabric treatment; catalytic domain; CD.

XX Acidothermus cellulolyticus.

XX US2004038334-A1.

XX 26-FEB-2004.

XX 28-JUL-2001; 2001US-00917376.

XX 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

XX (ADNE/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2004-203224/19.

PT Novel thermostable Aviii polypeptide of glycoside hydrolase family and

PT isolated from Acidothermus cellulolyticus, useful for degrading treated

PT biomass into simpler forms of carbohydrate.

XX Claim 28; SEQ ID NO 3; 19pp; English.

CC The invention relates to a thermostable cellulase enzyme, avicelase III

CC (Aviii) and its nucleic acid sequence. Aviii is useful for reducing

CC cellulose in a starting material. A thermostable Aviii peptide is useful

CC for degrading treated biomass into simpler forms of carbohydrate, which

CC is used in the formation of ethanol or other industrial chemicals. It is

CC also useful for treating fabrics to remove cellulose-containing stains.

CC The present sequence is Acidothermus cellulolyticus Aviii protein GHCA

CC catalytic domain (CD).

XX Sequence 733 AA;

Query Match 76.3%; Score 3920.5; DB 8; Length 733;
Best Local Similarity 99.1%; Pred. No. 4.6e-199;
Matches 733; Conservative 0; Mismatches 0; Indels 7; Gaps 7;

```
QY 47 ATTGPTWNSNVAIGGGGFVDGIVNEGAPGILYVRTDIGMYRMDANGRWIPLLDWGM 106
DB 1 ATTGPTWNSNVAIGGGGFVDGIVNEGAPGILYVRTDIGMYRMDANGRWIPLLDWGM 60
QY 107 NNMGYNGVSIADPINTNKVMAAVGMVYTNMWDNDGAILRSSDQATWQITPLPFLKGG 166
DB 61 NNMGYNGVSIADPINTNKVMAAVGMVYTNMWDNDGAILRSSDQATWQITPLPFLKGG 120
QY 167 NMPGRGGERLAVDPNNNDILYFGAPSGKLMRSTSGATWSQMTNFPDVGYIYANPTDT 226
DB 121 NMPGRGGERLAVDPNNNDILYFGAPSGKLMRSTSGATWSQMTNFPDVGYIYANPTDT 178
QY 227 TGYOSDIGVWVWAFDKSSSSSLGQASKTIFVGVADPNNPFWRSRDGATQAVPGATGF 286
DB 179 TGYOSDIGVWVWAFDKSSSSSLGQASKTIFVGVADPNNPFWRSRDGATQAVPGATGF 237
```


QY IPHKGVDPVNVHLYIATISNTGGPYDSSGDYVWFSPVSGTWTIRISPEVSTDIANDYFCY 346
 DB IPHK- FPDVNVHLYIATISNTGGPYDSSGDYVWFSPVSGTWTIRISPEVSTDIANDYFCY 235
 QY SGLTIDRHPNTIMVATQISWMPDIIIFSTDGATWTRIMWTSYPNRSRLRYVLDISAE 406
 DB SGLTIDRHPNTIMVATQISWMPDIIIFSTDGATWTRIMWTSYPNRSRLRYVLDISAE 355
 QY PWLTGVOQNPVPVPSFKLGMDRMAAIDPNSDRMLYGTGATLYATNDLTKMDSGQIHI 466
 DB PWLTGVOQNPVPVPSFKLGMDRMAAIDPNSDRMLYGTGATLYATNDLTKMDSGQIHI 415
 QY APWVGLEETAVNDLISPSGAPLISALGPGFTADTAVPSTIFTSVFTTGTSTVDY 526
 DB APWVGLEETAVNDLISPSGAPLISALGPGFTADTAVPSTIFTSVFTTGTSTVDY 474
 QY AELNPSIIVRAGSPDSSQPNDRHVAFSTDGKQWFGSGEPGVTTCGTVAASADGSRFV 586
 DB AELNPSIIVRAGSPDSSQPNDRHVAFSTDGKQWFGSGEPGVTTCGTVAASADGSRFV 534
 QY WAPGDPQGVYVAVGFGNSMAASQGVPAANAQISDRVNPFTFYALNSGTFRSTDGVTF 646
 DB WAPGDPQGVYVAVGFGNSMAASQGVPAANAQISDRVNPFTFYALNSGTFRSTDGVTF 594
 QY QVVAAGLPSSGAVGVWFHVPKSGDLMTLAASGLVHSTNGSSMSAITGVSSAVNVGFG 706
 DB QVVAAGLPSSGAVGVWFHVPKSGDLMTLAASGLVHSTNGSSMSAITGVSSAVNVGFG 654
 QY KSAFGSSYPAVFAVGTIGVTAQRSDDCGTTWVLTINDQHOYQNMCOALTGHANLRRV 766
 DB KSAFGSSYPAVFAVGTIGVTAQRSDDCGTTWVLTINDQHOYQNMCOALTGHANLRRV 714
 QY YIGTNGRGIVYGDIGAPSG 786
 DB YIGTNGRGIVYGDIGAPSG 713
 RESULT 8
 ABB99489
 ID ABB99489 standard; protein; 940 AA.
 AC ABB99489;
 DT 03-MAR-2003 (first entry)
 DE Amino acid sequence of a xyloglucanase enzyme.
 KW Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;
 KM textile scouring; cellulose fiber processing; ratting.
 OS *Jonesia* sp.
 FN WO200277242-A2.
 XX 03-OCT-2002.
 PD 27-MAR-2002; 2002WO-DK000210.
 PF 27-MAR-2001; 2001DK-00000504.
 PR (NOVO) NOVOZYMES AS.
 PA Duffner F, Sjoelom C;
 PI WPI; 2003-092855/08.
 DR N-PSDB; ABV76941.
 XX
 XX New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,
 PT and endogenous to a bacterium, useful in the textile industry for
 PT improving properties of cellulosic fibers, yarn or fabric.
 PS Claim 5; Page 72-75; 76pp; English.
 XX

CC The present sequence represents a xyloglucanase enzyme, belonging to
 CC family 74 of glycosyl hydrolases. The enzyme is isolated from *Jonesia* sp.
 CC DSM14140. The enzyme has been useful in processes for machine treatment of
 CC fabrics. It is also useful in the textile industry for improving the
 CC properties of cellulosic fibers, yarn, woven or non-woven fabric, and in
 CC a textile scouring process step. The xyloglucanase enzyme is also useful
 CC in the cellulose fiber processing industry for ratting of fibers such as
 CC hemp, jute, flax and linen. It is useful for preventing binding of
 CC certain soils to the xyloglucan left on the cellulosic material

Sequence 940 AA;

Query Match 42.3%; Score 2174; DB 6; Length 940;

Best Local Similarity 44.8%; Pred. No. 1.3e-106;

Matches 430; Conservative 141; Mismatches 311; Indels 78; Gaps 16;

QY TMSRRIVSLAATASFAVAALGVPLIATASAAHAATTPYTSNVAICGGGFVGV 69
 DB TMSRRIVSLAATASFAVAALGVPLIATASAAHAATTPYTSNVAICGGGFVGV 61
 QY TTVATATLALTATAGGIAATS--TPMPVAPSPSSPGYSMSNVEIVGGGFVGV 61
 DB TTVATATLALTATAGGIAATS--TPMPVAPSPSSPGYSMSNVEIVGGGFVGV 61
 QY FNEBAPGILVYRNDIGAYRMDANGRWIPLLDWGMNMGVGVSIADPIYTNKYWA 129
 DB FNEBAPGILVYRNDIGAYRMDANGRWIPLLDWGMNMGVGVSIADPIYTNKYWA 121
 QY FNGKDPGLVARTDIGAYRLNDSTGRNIPLDTHIGMDWSGILSLATDPVDTRVYL 121
 DB FNGKDPGLVARTDIGAYRLNDSTGRNIPLDTHIGMDWSGILSLATDPVDTRVYL 181
 QY AVGYTNSWPDNDGAILRSSDQATWQITPLPKLGMPKRGMERLAVDPNNDLTYF 189
 DB AVGYTNSWPDNDGAILRSSDQATWQITPLPKLGMPKRGMERLAVDPNNDLTYF 122
 QY AAGTYSMNDMPQONAILRSADKGETWETKMLPFRVGMMPKRGMERLAVDPNNKLYF 181
 DB AAGTYSMNDMPQONAILRSADKGETWETKMLPFRVGMMPKRGMERLAVDPNNKLYF 181
 QY GAPSCKGLMSTDSGATWSQMTNFPDVGTYIANPDTTG-YQSDIQQYVWVAFKSSSL 248
 DB GAPSCKGLMSTDSGATWSQMTNFPDVGTYIANPDTTG-YQSDIQQYVWVAFKSSSL 238
 QY GAGSNGMLKSTDYGTAKTGKVTSEFNAGNYVA---DASGAYTQGNQGVVWVTPPTSAKA 238
 DB GAGSNGMLKSTDYGTAKTGKVTSEFNAGNYVA---DASGAYTQGNQGVVWVTPPTSAKA 298
 QY GQASKTIFVGVADNNNVFWMSRDGATWQAPGPTGIPHKGVDPNVNHYATNTSG 308
 DB GQASKTIFVGVADNNNVFWMSRDGATWQAPGPTGIPHKGVDPNVNHYATNTSG 239
 QY GQTTQTYVGVADKQNNVYRSTDGATWQRPQPTGFLAQKQGHKQGLYATSDTG 298
 DB GQTTQTYVGVADKQNNVYRSTDGATWQRPQPTGFLAQKQGHKQGLYATSDTG 368
 QY GPYDSSGDVWKFVSTGWTIRISPEVSTDIANDYFCYSGTLTROHPNTIMVATQISW 368
 DB GPYDSSGDVWKFVSTGWTIRISPEVSTDIANDYFCYSGTLTROHPNTIMVATQISW 299
 QY GPYDSSKADVWRLDISGQWTRISPIBET-SSNAGFGSGLAIDRKNDITIMVSSQVSW 357
 DB GPYDSSKADVWRLDISGQWTRISPIBET-SSNAGFGSGLAIDRKNDITIMVSSQVSW 358
 QY PDMVYRSTDRGKTWSPIMELNGSQPRTKQNHVDYSGAPMLDFGNATKEPRANPKLGMWT 417
 DB PDMVYRSTDRGKTWSPIMELNGSQPRTKQNHVDYSGAPMLDFGNATKEPRANPKLGMWT 428
 QY PDIIFRSTDGATWTRIMWTSYPNRSRLRYVLDISAEPWLTGVOQNPVPVPSFKLGMD 428
 DB PDIIFRSTDGATWTRIMWTSYPNRSRLRYVLDISAEPWLTGVOQNPVPVPSFKLGMD 429
 QY EAMAIDPNSDRMLYGTGATLYATNDLTKMDSGQIHIAPMKGLERAVNDLISPSGA 488
 DB EAMAIDPNSDRMLYGTGATLYATNDLTKMDSGQIHIAPMKGLERAVNDLISPSGA 418
 QY QSPFIDPNSDRPFYGTGAGIYGTNLTNMKKGVKVDITYVAQGLEPAAQDLAAPPNI 477
 DB QSPFIDPNSDRPFYGTGAGIYGTNLTNMKKGVKVDITYVAQGLEPAAQDLAAPPNI 489
 QY PLISALGDLGFTADTAVPST-IFTSVFTTGTSTVDYAEINPSIIVRAGSPDSSQPN 547
 DB PLISALGDLGFTADTAVPST-IFTSVFTTGTSTVDYAEINPSIIVRAGSPDSSQPN 478
 QY GLYSALADIGFTHNDISQVFNKYYKNPHDITVSIIDPABEKPAVTVRAK--SISGET 535
 DB GLYSALADIGFTHNDISQVFNKYYKNPHDITVSIIDPABEKPAVTVRAK--SISGET 548
 QY DRHVAFTDGGKNNFQSGEPGVTTCGTVAASADGSRFVWAPGDPGVYVAVGFGNSMA 607
 DB DRHVAFTDGGKNNFQSGEPGVTTCGTVAASADGSRFVWAPGDPGVYVAVGFGNSMA 536
 QY TSWGVSTDDAGEITKPKATPSGVKPGSITVSAANSIYMAW--EAGAPRSTNGSSMS 593
 DB TSWGVSTDDAGEITKPKATPSGVKPGSITVSAANSIYMAW--EAGAPRSTNGSSMS 608
 QY ASQGVPAQAQIRSDRVNPKFTYALNSGTFRSTDGVTTFQVYA-AGLPSSGAVGMFHAV 666
 DB ASQGVPAQAQIRSDRVNPKFTYALNSGTFRSTDGVTTFQVYA-AGLPSSGAVGMFHAV 594
 QY TVSGILPHNAQVADRNVNANTLYGFVDGKFYHSTNGASFTASAFGFTPSG--NVRFRAV 651
 DB TVSGILPHNAQVADRNVNANTLYGFVDGKFYHSTNGASFTASAFGFTPSG--NVRFRAV 667
 QY PKEGEDLMA-----ASSGLVHSTNGSSMSAITGVSSAVNVGFGKSPAGSYPAVFAVVG 721
 DB PKEGEDLMA-----ASSGLVHSTNGSSMSAITGVSSAVNVGFGKSPAGSYPAVFAVVG 652
 QY PGRQGLHMLAGVSGSTYGMWRSTDGKNNKTVSGVQGDGAVGFKALASSGPIVITSA 711
 DB PGRQGLHMLAGVSGSTYGMWRSTDGKNNKTVSGVQGDGAVGFKALASSGPIVITSA 722
 QY TIGGVTAAYRSDDCGTTWVLTINDQHOYQNMCOALTGHANLRRVYIGTNGRGIVYGDIG 781
 DB TIGGVTAAYRSDDCGTTWVLTINDQHOYQNMCOALTGHANLRRVYIGTNGRGIVYGDIG 712
 QY KIDGVRGIFRSDDBKTKRINDNHNQWYMGASITGDPDVGVGRYIIGTNGRGIVYGDIG 769
 DB KIDGVRGIFRSDDBKTKRINDNHNQWYMGASITGDPDVGVGRYIIGTNGRGIVYGDIG 782
 QY GAPSGLSPSVSPSAPSLSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 840
 DB GAPSGLSPSVSPSAPSLSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 770
 QY -----SSTPPPGQDSGASGNTGATPTPPTPDSHNTGCG 804
 DB -----SSTPPPGQDSGASGNTGATPTPPTPDSHNTGCG 841
 QY SASPSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 900
 DB SASPSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 900

Db 805 STDGETSEBPGTGGSGGTRAPDSGNTSPGNTSSCKRYRSTTDMGSG---FTGAATINTSP 861
 QY 901 SSVDLSVTYVRY-----WFRD--GGSSLTLYNCDMAAIGCNIRASFGSVNPAIP 949
 Db 862 SPINQWTLAFTYPSGQTSSVMSATQTLSGRNVVLKNSGM-----NPTIP 906

RESULT 9

ADD22927
 ID ADD22927 standard; protein; 726 AA.
 XX
 AC ADD22927;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Aspergillus aculeatus avicelase III catalytic domain.

XX
 KM enzyme; Aviii; cellulose reduction; agricultural biomass;
 KM municipal solid waste; glycoside hydrolase; avicelase.
 XX
 OS Aspergillus aculeatus.

XX US2003108988-A1.

XX 12-JUN-2003.

XX 18-OCT-2002; 2002US-00155400.

XX 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

XX (ADNEY/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-810853/76.

XX New isolated thermal tolerant avicelase polynucleotide useful for

XX detection of a polynucleotide encoding Aviii and for reducing cellulose

XX in a starting material, e.g. municipal solid waste.

XX Example 2; SEQ ID NO 7; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a
 CC thermostable Aviii polypeptide. The polynucleotide is useful for
 CC detection of a polynucleotide encoding Aviii. The polynucleotide is
 CC useful for reducing cellulose in a starting material which involves
 CC administering to the starting material, e.g. agricultural biomass or
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
 CC method further comprises administering a second polypeptide molecule
 CC chosen from the glycoside hydrolase family of proteins. The present
 CC sequence represents the amino acid sequence of Aspergillus aculeatus
 CC avicelase III catalytic domain.

XX Sequence 726 AA;

XX Query Match 32.7%; Score 1680; DB 7; Length 726;

XX Best Local Similarity 46.2%; Pred. No. 1.3e-80;

XX Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 47 ATTGPWTGNVAI-GGGGFTDGIIVNEGAGIILYFRDIGMYRMDANGRWIPLDMWG 105

Db 1 AASQATYTKNVVTVTGGGGFTGPIVENPSAGVAAFTDIGAARLN-SDTWTPLMDWG 59

QY 106 ---MNNWYGVVSIADPINTNKVAAVGMATNSMDPNDGAILRSSDQATQITPLPF 162

Db 60 NDTWHDW---GIDALATDPVDTDRVYVAVGMATNEMDPNVGSILRSDQDITMTETLPLF 116

QY 163 KLGNNPGRGMRGLAVDPNDNIIYFGAPSGKLMRSTDSGATWSQMTNPPDVGYTIAN 222

Db 117 KVGNNPGRGMRGLAVDPNKNSILYFGARSGHGLMKLSTDYGATWSNVTSFTWTGYFOD 176
 QY 223 PRTTGTGSPDIOGVWVAPEKSSSILGASKITFVGVADNNPNVFWSRDGAATQAVGA 282
 Db 177 SSST--YTSDPVGAIAWTFEDSTSGSGSATPRLFVGADGKSVFSESDGATWAVSGE 234
 QY 283 PT-GFIPHKGVFPVNVHLIYATSNGTGPDGSGGVKPSVTSGTWTRISPVPSDTAN 341
 Db 235 PÖYGFIPHKGVLSPEERKTLIYSANGAGPYDGTNGYTHKNITISGWTIDSP---TSLAS 291
 QY 342 DYFGSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDTISYENRSLRYL 401
 Db 292 TYGYGGLSTVDLQVPGTAVAAALNCWPPDELIIFRSTDSGATWSPIMEMNGYPSINYYSY 351
 QY 402 DISAEPWLTTRGVQPNP-PVPSPKLGMMDEMALIDPFPSDMLYGTATLTAAINTLTQDS 460
 Db 352 DISNAPWIDTSTTDQFPV--RVGMVVEALALDPPDSNMLYGTGLTVGGHDLTWDS 408
 QY 461 GGGIHLAPWYKGLERTVNDLISPPSGAPLISALGDLGFTHADVTAVPSITIFSPPFTT 520
 Db 409 KHNVTYKSLAVGLEMAVGLITPPGSPALLSAVGDGFTYHSDLDAAFPQATHTPTTGT 468
 QY 521 GTSVDYALNPSTIIVRAGSFDPSSOPNDRHVAFTDGGKXWFOGSEPGVTTGTVASA 580
 Db 469 TNGIDYAGNKPSTNIVSAGASD--DYPT--LALSSNGSTWYADVASTGTGAVALSA 523
 QY 581 DGRFPWAPGDPCQPVVYAVFGNSWMAAGCPANAOIRSDRVNPKTFYALSNTFTFRST 640
 Db 524 DGDVTLLMSTSGALVSKSQG--TLTAAVSLPSGAIVADSKDNTVFYGGSAIAIYVSK 580
 QY 641 DGVTPQVPAAGLPSGAGVGMFHAIVKRGDMLLAASGLYHSTNGGSSWSAI--TVSS 639
 Db 581 NTATSPFKTVS-LGSSTTVNAI-RAHPSIAGDVWASTDKGLMSTDYGSTFTQIGSGVTA 638
 QY 700 AVNWFGKSAFGSSYPAPVFWGTIGVTGAYRSDCCCTTWWLINDDOHYGNMGA--IT 757
 Db 639 GMSGTFKASSTGSYVYIYGFITLDGAAGLFKSEDACTNMQVISAHAGFGS-GSANNVN 697
 QY 758 GDHANLIRVYITNGRGIYVDIGAPSG 786
 Db 698 GDLQTYGRVFRGHERPGLLRQSGREBAG 726

RESULT 10

ADD42055
 ID ADD42055 standard; protein; 838 AA.

XX ADD42055;

XX 15-JAN-2004 (first entry)

XX Trichoderma reesei Family 74 xyloglucanase.

XX Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;

XX xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;

XX ethanol production; detergent composition; fabric treatment;

XX textile treatment; enzyme.

XX Hypocrea jecorina.

XX WO2003089598-A2.

XX 30-OCT-2003.

XX 17-APR-2003; 2003WO-US011831.

XX 19-APR-2002; 2002US-0373987P.

XX (NOVO) NOVOZYMES BIOTECH INC.

XX Michael R, Zaretsky E, Haas J;

XX WPI; 2003-845528/78.

DR N-PSDB; ADD42054, ADD42060.

XX New polypeptides having Family 74 xyloglucanase activity, and encoding
PT nucleic acid molecules, useful for degrading cellulose- and hemicellulose
PT -containing biomasses to ethanol or as a detergent.

PS Claim 8; SEQ ID NO 2; 96pp; English.

XX The invention relates to a Family 74 xyloglucanase (Add42055) from the
CC fungus *Trichoderma reesei* (Hypocrea jecorina), and nucleic acids encoding
CC it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic
CC linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.
CC The invention also relates polypeptide sequences at least 70% identical
CC to the enzyme, expression vectors and host cells comprising a nucleic
CC acid of the invention, the recombinant production of the enzyme, and
CC mutant enzymes and the nucleic acids encoding them. The xyloglucanase of
CC the invention can be used in the degradation of cellulose- and
CC hemicellulose-containing biomasses to produce ethanol. It can also be used
CC in a detergent composition for treating fabric during a machine washing
CC cycle. The present sequence represents the *Trichoderma reesei* Family 74
CC xyloglucanase.

XX Sequence 838 AA;

Query Match 31.7%; Score 1625.5; DB 7; Length 838;
Best Local Similarity 41.7%; Pred. No. 1.1e-77;
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;

25 SFVAALGLVLAITATSPAAHAATTOPTYSNVAI-GGGFVGVGIVNEGAPGLLYVRTD 83

4 SRLVALVLAIV-----IPAHAA-----FSMKVNLGLGGGGFVPEIIPHPTKGVAVARTD 53

84 IGGMYRMDAANGEMIPLLDV-----GNNNGYGVSIADDPINTKVMVAVMYTNMSWD 139

54 IGLLYRLN-ADDSMTAVTDGIDADNAGWHNN-----GIDVALDPODDQKRYAVAGMYTNMSWD 109

140 PUNGALIRSSDQATQITPLPKLGNNMGRGGEELADPNNDNLIFGAPSGKGLMR 199

110 PMSGALIRSSDRGATWSFTNLPRKVGNNMGRGGEELADVPANSNITIFGAPSGKGLMR 169

200 STDSGATWSQMTNPPDVGTIANPTDTTGYOSDIOGVVWVAFDCKSSSLGQASRTIIFGV 259

170 STDGGVTFKSVSSFTATGTTIIPBDSNGNSDKGLMWTFTBDSSTTGATSRIFVGT 229

260 ADP-NNPVFWSRDGATWQAVPGAFTGFIPIHKGVDFVNHVLYATSNLTGGPYDSSGDV 318

230 ADNITASVYVSTNAGSTWASVAPGQPKYFPHKAKLQPAEKALYLTSYSGTGPYDGLTSV 289

319 WKSSTVSGTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIMATQISWPDITIFRSTD 378

290 WRVDIAGGTWKDITPVSGSDI--YFGFGGLGLDQKRGTLVVASLNSWPDADLFRSTD 346

379 GGAITWRIMDTSYSPNRSLLRYVLDISABPMLTFG-VQPNPVPVS-----PKGMDENAMAI 433

347 SGTWWSPIWMASTPTEFTYYISITPKAPWIKNNFIDVTSRSPDGLIKRIGMIESLEI 406

434 DPFNSDRLXGTGATLYATNDLTKMDSGQIHIAPWYKGELETVNDLISPPSGAPLISA 493

407 DPFDSNNMWTGTTIRGCHDLTWMDRHNVSIGSLADGIEBFVODLASAPGSELLAA 466

494 LGDLGFTHA--DVTAVPSTIFTSPVFTTGSVDYVLAELNSIIVRAGSPSSSQPNDRH 550

467 VGDNGFTFASRNDLGTSPQVMAVPTWATSTVDYAGNSVKSIVRVAGNTAGTQO----- 521

551 VASTDGGKWFQSGSEGGVTTGGTVASADGSRFWAPGPGQPVYAVAFGNSMASQ 610

522 VAISSDGATWSIDYAADTSMNGGTIVYASDGTILWSTASSG--VQRSGFQSGFASVS 578

611 GVPANQIRSDRVNPKTFYVALLSNGTFRSTGGVTFQVVAAGLPSSGAVGVW--FHAVPG 668

579 SLPRGAVIASDKTNSVFIYAGSGSTFYYSKDTGSSF--TRG-PKLSAGTIRDAAMP 634

669 KEGDLMLAASGLYHSTNGGSWSGAI-TGVASVAVNVGFKGAPSSSYPAVFAVGTIGV 727

DB 635 TAGTLVYSTVGVGIFRSTDSGTFEQGVSTALNTYQIALGVGS-GSNW-NLYAFGT--GPS 690

QY 728 GA--YRSDDCGTTWVLLINDDOHQYGNWGAITGPHANILRVYIGTNGRGIVY--GDIGGA 783

DB 691 GARLYASDGSASTWTDIQSGQSPSISTKYAGSGSTAGVYVYTNRGVFIYAGTYGGG 750

QY 784 PSGSPSPSVSPSAPSAS 843

DB 751 TCGT-----SSSTKQSSSTSSASSSTTLKSSVSTTRASTVSSRTSSAAGPTGS 801

RESULT 11
ADH51581
ID ADH51581 standard; protein; 818 AA.

AC ADH51581;

DT 25-MAR-2004 (first entry)

DE Trichoderma reesei EGVI protein sequence related to ethanol production.

KW EGVI protein; endoglucanase; ethanol production; biomass composition;

KW fermentation; cellulohydrolase; detergent production; softening agent;

XX cotton fabric; wood pulp degradation; sugar; enzyme.

XX Hypocrea jecorina.

XX Key Location/Qualifiers

FT Misc-difference 1..818

FT /label= OTHER

FT /note= "OTHER= All Xaa's given in this sequence are

FT unknown amino acids, given as J in the specification.

FT They should probably be Ile."

XX US2003113732-A1.

XX 19-JUN-2003.

XX 18-DEC-2001; 2001US-00026994.

XX 18-DEC-2001; 2001US-00026994.

XX (DUNN/) DUNN-COLEMAN N.

XX (GOED/) GOEDGEBOUR F.

XX (WARD/) WARD J.

XX (YAOJ/) YAO J.

XX Dunn-Coleman N, Goedgebuer F, Ward M, Yao J;

XX WPI; 2004-106460/11.

XX N-PSDB; ADH51580, ADH51583.

XX New substantially purified EGVI polypeptide with the biological activity

XX of endoglucanase, useful for degrading biomasses to ethanol.

XX Claim 18; SEQ ID NO 2; 26pp; English.

XX This invention relates to a novel purified EGVI protein with the
XX biological activity of an endoglucanase, comprising an amino acid
XX sequence chosen from a fully defined sequence as given in specification.
XX The protein is useful for producing ethanol, which involves contacting a
XX biomass composition with an enzymatic composition containing the protein
XX of the invention to yield a sugar solution, adding to the sugar solution
XX a fermentative microorganism, and culturing the fermentative
XX microorganism under conditions sufficient to produce ethanol, where the
XX biomass composition may be optionally pretreated. The method further
XX involves the addition of at least one endoglucanase or cellulohydrolase.
XX The pretreatment is with a dilute acid. The protein of the invention is
XX also useful for the production of a detergent composition which may be
XX used as softening agent and for improving the feel of cotton fabrics and
XX for degrading wood pulp into sugars. The present sequence is that of the
XX T reesei EGVI protein of the invention.


```
XX Sequence 818 AA;
SQ Query Match 30.9%; Score 1588; DB 8; Length 818;
Best Local Similarity 41.6%; Pred. No. 1.1e-75;
Matches 338; Conservative 126; Mismatches 295; Indels 54; Gaps 21;

OY 52 YTSNVAI GGGGFDGIVFNEGAGIIVRTDIGMTRMDANGWIPILDMV---GW 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 FSWKNVKLGCGGFPVGLIFHPKTKGVAARFDIGLYRLN-ADDSMTAVTDGIADNAGW 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 107 NNMGNGVNIADPNTNKWAAVGMATNSMDPNDAIRSSDOGAQWQITPLPKLGG 166
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 HNW---GIDAVALDPDDQKVTAAGVATNSMDPSGALIRSSDRGATMSFTLPRKVG 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 167 NMPGRGMBRLAVDPNNNDNIIYFGAPSGKGLMRSTDSGATWSQMTNPDVGIYIANPTDT 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 118 NMPGRGMBRLAVDPNNNSNIYFGARSGNGLMKSTDGVTFSKVSFTATGYXPDPSDS 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 227 TGYSDIGQVAVAPDKSSSLGQASKTIFGVADP-NNPVFSRDCGATWQAVPGAPTG 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 NGVNSDKQGMWVTFPSTSTTGATSRIFVGTADNITASVYVSTAGSWSAVPQPGK 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 286 FIPHGVPDPVNHVLIATSNITGPDSSGDVWKFSTVSGTTRISPVSTDPADYFG 345
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 238 YFPHAKIQPAEKALYLYSDGTGPDYDGLGSWMRYDLAGIWKDXTPVSGDL---YFG 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 346 YSGLTIDRQHPNTIMVATOISWMPDTIIFRSTDCGATWTRIMWTSYPNKSLRYVLDISA 405
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 225 FCGLGIDLOKPGTLVVASLNSWMPDQOLFRSTDSGTTWSIMWASPTETYYISISTPK 354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 406 EFWLTFG-VQNPVPVS---PFLGWMDEMAIDPNSDBMLYGTGATLYATNDLTKWDS 460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 355 APWIKNNFIDVISESPSDGLIKRIGMMWISLEIDPDSNMLYGTGWTGKGHDLITWMDT 414
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 461 GGOIHAPVWKGLEFAVNDLISPPSGAPLISALDGLGFTHA---DVTAPSTIPTSVP 517
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 415 RHNVSIQSLADGIEEFSVODLASAPGSEILAAVANGFTTFASRNDLGTSPQVWMAPT 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 518 FTTGTSVDYAEINPSIIVRAGSFPDSQPNDRHVAESTDCAKMFQGSSEPGVTTGGTVA 577
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 475 WATISVVDYAGNSVKSIVRGN-----TAGTQVALISSDGAATWSIDYADTSMNGTVA 528
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 578 ASADGSRFYWAPDPCQPVVYAVGFGNSMAASQGVPAQAQIRSDRVNPKTFYALSNGTFY 637
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 529 YSADGPTILMSTASG---VQRSQFQGSFASVSLIPAGAVIADSKTNSVFYAGSGSTFY 585
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 638 RSTDGVTIQQVAAAGLPPSGAVGM--FHAVPQKEGDLMLAASSGLYHSTNGSSWGAI- 694
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 586 VKSDTGSF---TRG-PKIGSAGTIIPDIAHPTTAGTLVYSTDVGIIFRSTDSGTTFGQVS 641
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 695 TGVSASAVNIGFGSAGSSYPFAVFVGTIGAVTGA--YRSDCGTWTVLINDDOHQGNW 752
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 642 TALTNTYQIALGVGS-GSNW-NLYAFGT--GPGSARLYASGDSGASTTIDIGSGSFGSID 697
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 753 GOAITGDHANLRRVYIGTNGRGIVY--GDIGAPSGSPSPSVPSASPSLSPSPSSSP 810
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 698 STKVASSGTAAGVYVGTNGRGVFAAGTVGGGTGT-----SSSTQSSSSSTSSA 748
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 811 SPSPSPSSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 843
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 749 SSSSTLRSSVSTTRASTVTTSRTSSAAGPTGS 781
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
ADH19118
ID ADH19118 standard; protein; 812 AA.
XX
AC ADH19118;
XX
DT 11-MAR-2004 (first entry)
XX
DE Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme pre protein.
```

```
XX
KM xyloglucan oligosaccharide degradation; enzyme; genetic engineering;
KW precursor.
XX
OS Geotrichum sp. M128.
XX
PN EP1350844-A2.
XX
PD 08-OCT-2003.
XX
PF 25-MAR-2003; 2003EP-00251866.
XX
PR 25-MAR-2002; 2002JP-00083433.
XX
PA (NRAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX
PI Yaoli K, Mitsunishi Y;
XX
XX WPI: 2004-100948/11.
XX
DR N-PSDB; ADH19117.
XX
PT Novel xyloglucan oligosaccharide-degrading enzyme with a different
XX degradation mechanism from known enzymes.
XX
PS Claim 5; SEQ ID NO 12; 39pp; English.
XX
CC The invention relates to a novel xyloglucan oligosaccharide-degrading
CC enzyme with a different degradation mechanism from known enzymes. The
CC primary structure and polynucleotide structure of the xyloglucan
CC oligosaccharide-degrading enzyme provided by the present invention allows
CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation
CC activity to be prepared at a low cost through a genetic engineering
CC process. The current sequence is that of the Geotrichum sp. M128
CC xyloglucan oligosaccharide-degrading enzyme precursor protein of the
CC invention.
XX
SQ Sequence 812 AA;

Query Match 21.9%; Score 1125; DB 8; Length 812;
Best Local Similarity 34.5%; Pred. No. 3.2e-51;
Matches 285; Conservative 129; Mismatches 291; Indels 120; Gaps 31;

OY 31 ALGVLPPIAITASPAHAATTPQYTWSNVAIGGGGQVFDVFNEGAPGILYVRTDIGMTRW 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6 SLGALFALSLIASLAIAKEHYEFKNVAIGGGYITGIVAHPTKDLIYARTIDIGAYRW 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 91 DAANGRWIPLLDWGMNMGVNVVSIADPINTNKWAAVAGMT-NSWDPNDGAILRSS 149
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 66 DAGTSKILPLNDFIEADPMIMGTESIADPNNDRLYLQGRVYGDW---AAYFVSE 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 150 DQGAWQITPLPFLKLGMPGKMBGMBRLAVDPNNNDNIIYFGAPSGKGLMRSTDSGATWSQ 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 122 DRQGSFTIYESPFPMGANIDMGRNNGERLAVNPFSNENVMGTRT-EGIMKSSDRAKTWTN 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 210 MTNFPDVGTIYIANPTDTTIGYSDIGVYVWAFDSSSSSLGQASKTIFGVADPNNVPWS 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 181 VTSIPDAF-----TNGIGYTS-----VIFDP-----BRNGTISGATAPOG-MVYT 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 270 RDGATWQAVPGATPGFI-----PHKGVFPDPVNHVLIATSNITGAPY 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 221 HDGVSMBPVAAGQSSSWLNRTTGAFPDKKASIAQPHKVALTP--NFLVYTVADYPPPW 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 312 DGSSGDVWKSFTVSGTWTRI-----SPVSTDTANDYFGYSGLTIDRQHPNTIMVATQ 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 279 GVTGGEVWRQRTSGAMNDITPRVNSSPAPYNNQTFPAGFCGHSVADATPNNRLVIT- 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 365 ISWMPDTI---IFRSTDCGATWTRIMWTS-----YNNRSIRYVLDISABFWLTFG 412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 338 LDRDPGALDSIYISTAGATWQVTOQLSSPSNLGKNGMCHPTNARY-KDGTVPVWLPFN 396
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 413 VQP-----NRPVPSP---KLGWMDMAAIDPNSDRMLYGTGATLYATNDLTK-----WDS 460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 397 NGPQWGGYGAHGHGTGLTKFGWMSAVLIDFPNPHLMYGTGATITWADTILSRVAKDM-- 454
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```



```

QY 461 GQQLHIAF-----MVKGEETAYNDLISPPSGAPLISALDGGFTHADYVAVSTLTSP 51.6
Db 455 -----APSWYQIDIGIEENALLSLRSPSGAALLSGIDISMKRDLTK-DQKMGAP 507.7
QY 51.7 VFTTGTSVDYAEINPSIIIVRAGSPDPSQPNDRHVAVFSTDGQKMW--FQSEEGVYT-- 571.2
Db 508 QPSNIDSDIDAAGNFFVNVVRAASSSGHEIYDSACARGAYATDGDGMATIIFFICPPGMAASHY 56.7
QY 57.3 -GGTVAASADGSRFWAP--GDPGQPVVYAVGFGNSMAASQGVPA-----NAQIRSDRV 62.3
Db 568 QGSTIAVNDASSGQIYWSTLDEQASGPMWYSHDYGKTWS-----VPADGIKAQNTAVYLSDKV 62.3
QY 62.4 NPKTFYVLSNGTFPYNSTDGCVTTPQPYAAGLPSSGAVGVNFHAFVKEGDIWL--AASSGLY 66.2
Db 62.4 QDGTFFYATDGGKRFVSTDGGKSYAAKAGALVLT--GSLMPAVNPMWAGDVMVFVEPEGGLF 68.1
QY 66.2 HSTNGGSSMSAL-TGVSSAVNNGFGKS-----AGSSYPAVFVVG--IGGVTAARYSDDC 73.5
Db 68.2 HSTDEGASFTFRGTANATLVSGAPSKSKDKKKAASPAVFIWGTDPKSSDGLYRSDDN 74.1
QY 73.6 GTTWTWLINDQHOYGNMGOAITGDHNLNRRVYIGTAGRGIVYCDI 78.0
Db 74.2 GSTWTRVNDQEHNTSG-PTMIADPRTYGRVILGTINGRGIVYADL 78.5

```

RESULT 13

ID ADH19120 standard; protein; 789 AA.

ADH19120; AC

DT 11-MAR-2004 (first entry)

DE Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme mat protein.

KW xyloglucan oligosaccharide degradation; enzyme; genetic engineering;

XX

OS Geotrichum sp. M128.

PN EP1350844-A2.

PD 08-OCT-2003

PF 25-MAR-2003; 2003EP-00251866.

PR 25-MAR-2002; 2002JP-00083433.

PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY

PI Yaoi K, Mitsuishi Y;

DR WPI; 2004-100948/11.

XX

PT degradation mechanism from known enzymes.

PS Claim 2; SEQ ID NO 14; 39pp; English

CC The invention relates to a novel xyloglucan oligosaccharide-degrading

primary structure and polynucleotide structure of the xyloglucan

high-purity polypeptides having a xyloglucan oligosaccharide-degradation

CC process. The current sequence is that of the Geotrichum sp. M128

CC BY-NC-ND 4.0 International license.

Sequence 789 AA;
SQ

Query Match	21.8%;	Score 1117;	DB 8;	Length 789;
Best Local Similarity	35.0%;	Pred. No. 8.2e-51;		
Matches 281;	Conservative 124;	Mismatches 279;	Indels 120;	Gaps 31

Qy	52	ITWSVAIIGGGFVDGIVFENAGAPOLLYVRDIDGMYRMDANGRMIFLLPMVGNMNGY	111
Db	4	YEFKVAIAGGGGYITGIVAHHPKTKOLLARDDIGAVYMDGTSKMFIPLNDIFIAEQDNTI	63
Qy	112	NGVSIADPINTKVMVAAGMYT--NSWDPDGAIIIRSSDGA2TQWITPLPEKLGANNPG	170
Db	64	MGTEIADDPNNPRLIYLAQGRYVEDM---AATVSESDGQSPTIYESPPMGANDWG	119
Qy	171	RGMERLAVDPNDNITLYFGAPSGKGLRSTDSGATSCQMTNPFVGYTIANPTDTTGYQ	230
Db	120	RNNGERLAVNPFNSNEVMWGTRT--BGIKWSSPRAKTWNIVTISIPAF-----TNGIGYT	172
Qy	231	SDIQGVWVAPDKSSSLGQASKITFVGVADDPNNNVFVMSDGA2TQWQAVGAPGTGI---	287
Db	173	S-----VIDP-----ERNGTIYAATAPOG--MYTHDGQVSEMEPIAAGPSSMLART	218
Qy	288	-----PHKGVDPVNHVLYIATSNGTGPEYDSSGDVWKFSVTSGTWTRI--	331
Db	219	TGAFPDKKPASIAPQPMKVALTP--NPLYIYVADYPCGPMGVTFGEVWRQNRTSGAMNDIT	276
Qy	332	-----SEVPSTIDTANDYFCYSGLIIDQHPNTIWAQIISWMPPTI---1FRSDGAT	382
Db	277	PRVGNSSAPFVNNQTFPAGGFCGLSDVATNPRRLVIT--LDRDPGALDSTIYSTDAGT	335
Qy	383	WTRIMDMS-----YPNRSLRYVLDISAEPWLTFGVOP-----NPPVSP--KLG	425
Db	336	MKDVIQLSSPSNLBEGMWHPTNAARY--KDGTFVPLDFFNNPQWQGGYGAPOHPTGLTKFG	394
Qy	426	WMDERMAIDPENSDMLYGTGATLYAINDLK---WDSGOIHIAP---MVKGLEEYA	477
Db	395	WMMSYVLLIDPFEBHIMYGTGATIWATTLSEVERKW-----APSWYLOIDIIEENA	446
Qy	478	VNDLISPSGAPLISALDGLGFTHADYTAVASTFTSPFVPTGTSVDAELNPSIIVRA	537
Db	447	ITLSLSPSGAALLSGIDIGISMKRDDLK--POKFGAPQFSNLSIDAAGFPVAVVRA	505
Qy	538	GSFDRSSOPNDRHVAFFSTDGKNW--FOGSEBGVTT---GGTVASAADGSBFVAP--G	590
Db	506	GSSGHEYDSACARGAVATDGGDAWITPFCPFGMNAHSYQGSTIANVADSGSQTIVMSTKID	565
Qy	591	DPGQPVYTAVGFNSWMAASQGVPA-----NAQISDRNPKTFYALNSGTYYRSTDGAV	644
Db	566	EQASGPMYSHDGKJTWMS---VPADDLKQATQANTVLSDKQODTFYATDGGKFFSVTDGK	621
Qy	645	TFOQVPAAGLPSGAGVGMFHAVPKRGKDLML--AASSGLYHSTNGSGSMGAI--TGVSAAVN	702
Db	622	STAAKAGAGLVTT--GTSIMPAVNPWYAGDVWVPEBGLFHSHTDFGASFPRTVGTAAVTIYS	679
Qy	703	VGFQGS---APGSSYPAVFVVGTT--IGGVTAAYRSDCGTITVLLINDDOHOYGMWQAI	756
Db	680	VGAPSKSDQKKAASAPSAVFIWGTDKPGSDIGLYRSDDNGSTWTRVNDQEHNSYG--PTMI	738
Qy	757	TGDHANLRARYIGTNGRGIVYGD I 780	
Db	739	EADPKYGVYIGTNGRGIVADL 762	

RESULT 14

ID ADH19124 standard; protein; 826 AA.

ADH19124; AC

DT 11-MAR-2004 (first entry)

DE Geotrichum sp xyloglucan oligosacch-enzyme-derived protein

KW xyloglucan oligosaccharide degradation; enzyme; genetic engineering;

133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1

role in the differentiation of a plant cell. The current sequence is that of the Geotrichum sp. M128 xyloglucanendohydrolase precursor protein of the invention.

Sequence 776 AA;

Query Match 19.3%; Score 993; DB 8; Length 776;

Best Local Similarity 32.4%; Pred. No. 2,9e-44; Matches 273; Conservative 127; Mismatches 289; Indels 154; Gaps 34;

```
QY 11 MRSRLVSLAATASFAVAALGVLPATASPAHAATQPYTWSNVAIGGGFVDGIVF 70
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 LRTFGLALVCAVGSWAATVABL-----KPYTISGGGFISGLVA 41

QY 71 NEGAPGLVPTDIDGATWMDAANGWITPLDWN-----GMNNGYNGVSIADPINTN 125
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
42 HPTKDLIYARDIDGTYRWNAKMEWETPTDFIINNALANGANLLGTESIALDPNPD 101

QY 126 KVMVAVGMWYNSMDPNDAILRSSDQATWQITPLPKLGNNMGRGMGERLAVDPNND 185
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
102 RLYLAQGDIV-QMDP-MAAFIVSDDRGKTFKQYRSPVPMGANDMGRNGERLAVNPHWT 159

QY 186 ILYPGAPSGKGLMRSTSGATWSQMTNFPDVGTYIANPTDTGYQSDIQGVWVAFDKSS 245
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
160 ELMPGSRIT-QGLMRSTDRAGTWSRMNQLPDSSY-----GI-----GISVIFDPKN 205

QY 246 SSLGQASKTIFVGVADPNNPVFWSRDQATWQAVPGAPF-----G 285
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
206 VGTAVYASHAVAGLW-----VTW--DGANMSQVGGQPTQMSDWTKSIVAASGTALQSSG 258

QY 286 FIPHKGVFPVNVHLYIATSNNGPYDGSQDVWPKPSVTSGTWRISFVPSTDFANDY-- 343
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
259 PLPIKIALGK-NGRLITYSDAPGWPGLYGEVWSYDPTGNMKHIT--PSRGAANTYPA 315

QY 344 -----FGYSGLTIDROHPNTIMATQISWMPDTIIFRSTDGATWTRIMDTSYPN 394
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
316 PTGNKKVVGGMNGISVG--NGDTVVYSTIDANGEDS-VLSLRDAGNSWKDLGKLTTPAG 372

QY 395 RSLRYVLDISAE-----PWLTFGVQPNPVPSPKLGMDAMALDPFNSDRMLYGTGA 447
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
373 AGNSQKESDADKLRNGTPLPLWLSFQNRSGIV--GFGMWLAAILLDPF-SDRLLYGTGA 428

QY 448 TLVATDILTKMDSGGQIHIAFM-----VKGLETAVNDLISPPSG-APLISALGDLGFTH 502
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
429 VIMATVAVSRADS---NOAPSWYINTEGIEETAIIVLKSPAPPAHLFGMYDLGGWRH 484

QY 503 ADVTAVPSTIFPVTFTTGSVDYAEINPGLIIVRAGSFD--PSSQPNDRHVAFSTDGKN 560
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
485 DDF-SVPOPMYSKPTSTSDGLDPAGRAANVLARVGRNDHPDAGVACTQGA YTTNSGDS 543

QY 561 W--FQSEBPG-GVTTGTVAAASADGRFVAAPGD-PGQPVYAVAFGNSWMAASQVPAHA 616
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
544 WTLFQTCVPSLEVGNGSTIAGADKTFVWSPSKADGKGPYTSDDYGTWTAAPSGLSKQT 603

QY 617 Q-IRSDRVNPKTFYALNSGTFYRSTGCVTFQPVAAGLPSSGAVGMFAVPG---KEG 671
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
604 TGIAADRVQANTFYVVEGDFPFVSTDGKSKYTKKNGLPCCWT---YTGTPVTSNLRAG 659

QY 672 DLMFLAAS-GLYHSTNGSSMSALITGVSAVNVGFGKAPSSYPAYFVVG----- 722
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
660 ELWVSVKGVGIYHSTDPGNTFTALAGSSSLN-----PAVFSIGAPQTPNATE 707

QY 723 ---IGGV-----TGAVRSDCGTWWLINDDQHQYGNWCAITGDHANLRRVYIGTNGR 773
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
708 TLFKNGIPSAQPEGLYMSTDNGMLTRLNDDAHNYGG-ATVISGDPRIYGRVYI GMMGR 766

QY 774 GIV 776
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
767 GII 769
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:48:13 ; Search time 26.555 Seconds
(without alignments)
3467.504 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRLTKMSRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: dirl:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2436	47.4	890	2 T35237	probable secreted
2	2037	39.7	839	2 D97013	Avicelase III - As
3	1740	33.9	856	2 T00349	hypothetical prote
4	1161	22.6	707	2 F72393	xylinase - Caldice
5	431.5	8.4	1779	2 T31085	1,4-beta-glucanase
6	398.5	7.8	1711	2 T31337	cellulase (EC 3.2.
7	397.5	7.7	1742	2 T17120	hypothetical prote
8	372.5	7.3	473	2 S50755	salivase - Actino
9	346.5	6.7	901	2 A49227	mannan endo-1,4-be
10	340.5	6.6	1331	2 A48954	hypothetical prote
11	308.5	6.0	611	2 S76211	hypothetical prote
12	301.5	5.9	1749	2 S75138	hypothetical prote
13	293	5.7	279	2 T10361	hypothetical prote
14	291	5.7	351	2 S50754	hypothetical prote
15	290.5	5.7	2232	2 T34434	hypothetical prote
16	290	5.6	1039	2 S02711	cellulase (EC 3.2.
17	288.5	5.6	1032	2 T34433	hypothetical prote
18	283.5	5.5	474	2 S15921	protein TPX-VT3 -
19	262	5.1	2468	2 A83412	hypothetical prote
20	261.5	5.1	1854	2 S36859	cDPA protein - Clo
21	254.5	5.0	913	2 S20580	exo-alpha-sialidas
22	246	4.8	915	2 A43802	cellulase (EC 3.2.
23	245	4.8	879	2 A47704	endoglucanase I (E
24	244	4.8	1664	2 T18262	S-layer protein -
25	243.5	4.7	552	2 T08148	proline-rich myos
26	243	4.7	288	2 T17737	proline-rich prote
27	235.5	4.6	2124	2 A28452	proteoglycan core
28	229.5	4.5	496	2 T17908	proline/lysine-ric
29	227.5	4.4	2271	2 F90073	hypothetical prote

30	222.5	4.3	360	2 S12850	protein TPX - Ther
31	222.5	4.3	3570	2 T45025	mucin MUC5B, trach
32	221	4.3	2132	1 A55182	aggregan precursor
33	219	4.3	4776	2 E95206	cell wall surface
34	218.5	4.3	3164	1 WMBEH6	U36 protein - hum
35	217	4.2	241	2 AC2284	hypothetical prote
36	216	4.2	383	2 T46707	proteophosphoglyc
37	216	4.2	449	2 S16748	hypothetical prote
38	215.5	4.2	13055	2 T16580	hypothetical prote
39	215	4.2	339	2 T17636	proline-rich prote
40	215	4.2	13288	2 T03099	mucin, submaxillar
41	214	4.2	300	2 T02220	hydroxyproline-ric
42	214	4.2	532	2 B35621	spore germination
43	211.5	4.1	1275	2 T13369	hypothetical prote
44	211	4.1	602	2 AD2067	hypothetical prote
45	210	4.1	549	2 T17525	proline-rich prote

ALIGNMENTS

RESULT 1
T35237
probable secreted cellulase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: T35237
R:Seeger, K.J., Harris, D., Parkhill, J., Barrell, B.G., Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: 221572
A:Accession: T35237
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-890 <SE>
A:Cross-references: UNIPROT:086727; EMBL:AL031515; PTDN:CAA20642.1; GSPDB:GN00070; SCOE
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOE:DB:SC5C7.30C

Query Match 47.4%; Score 2436; DB 2; Length 890;
Best Local Similarity 48.6%; Pred. No. 2.3e-101;
Matches 475; Conservative 129; Mismatches 248; Indels 126; Gaps 19;

QY	11	MSRRLVSLAATSPFAVAALGVLPITATSPA--HAATQPTWNSVAIGGGFVDCI	68
DB	1	MRRTILTVLALAAGLA--GSPPAASAPPAVAADSYTKAKRIDGGFVDCI	57
QY	69	VFNKGAPGILVYRTDIGMYTDAAANGRWIPLLDWGNMNGVNSIAADPINTKRW	128
DB	58	VFNREKDLAARTDIGAYRWQESHMTPLDHDVGMDDGHTGVVALASDAVDPDRY	117
QY	129	AAVGMTNSWDPNDGAILRSSDQATWQITPLPFLGGMPCRGKGERLAVDPPNDNLY	188
DB	118	AAVGTYTNDMPDPTNGAVLRSDARGASWEKADLPFLGGMPCRGKGERLAVDPPNDVLY	177
QY	169	FGAPSGKLMSTDSGATWMSQMTNFPDYGTIYANPTDTGYSDIOGVVVAFDKSS--SS	247
DB	178	LGAPSGHLMSTDSGATWMSQMTNFPDYGTIYANPTDTGYSDIOGVVVAFDKSS--SS	237
QY	248	LGQAASKTIFVGVADPNPNPFWNSRDGATWQAVGAPPTGFI PKAGVDPVNHLYATSN	307
DB	238	AGTATRTLYGVADKEMNAVYRSTDGATWERAAGPTGILAKGVDAENGVLAYSDT	297
QY	308	GGPYDSSGDVWKFVSTGVTWRISPVSTDTANDYFGYSGLTIDRQHPNTIMVATQISW	367
DB	298	GGPYDGGKGLRVATATGWTDISPAABADT---YGFSGLTVDQRGGVMAATAYSSW	354
QY	368	WPDITIFSTDSGATWRTIMWTSTYPNLSLRVLDISAEPWLTFGVQNPVPSPRLGMM	427
DB	355	WPDITIFSTDSGATWQMSYTSYDRENRYTMVSSSPMLTGWANPAPPGTGLGMM	414
QY	428	DEAMALDPNSDRMLYGCATLYATNDLTKW--DSGGQIHIAVMYKLEEFANVDLISPS	486

Db 415 TEALEIDPPDSRMWYGTATVYGTENLTNMDDEGGTFAVEPMVRGLBETAVNDLASPPS 474
 Qy 487 GAPLISALGDLGGFTHADVTAVPSTIFTSPTVFTTGTSDVYAEINPSIIYRAGSPDPSQP 546
 Db 475 GAPLISALGDLGGFTSTLEVPVSMWYTSNFTSTSLDPAETKRPDVVAVAGNLD--SGP 532
 Qy 547 NDRHVAFSDGGKNNWQSGSEPGSVTTGTVAAASADSFRFVWAPGDQGPVVYVWFGNSW 606
 Db 533 ---HIAFTSDNANWVGTDPSGVSGGTVAAGADDSRFVWSP--EGAGVQYTTGFGTSM 587
 Qy 607 AASQGVANAOIRSDRVNPKTFYALSGNFTYRSTDDGVTPEQVAA--GLPSSGAVWVFHA 665
 Db 588 QASTLIPAGAIYESDVNNTATFYGFKSGRFYVSTDDGATFTASAATLPAAGD--GVAFKA 645
 Qy 666 VPKGEGDLMLAAS-----GLYHSTNGSSWSAITGVSSAVNVGFKGAPGSSYPAVFV 720
 Db 646 LFGGEGDVLWAGGAADGPGYGLMHSTDDGGFTRLPGVDADATVGFGRKAPGASQYTLFTS 705
 Qy 721 GTTGGVTGAYRSDDCCTTWVTLINDDOHYGNMGQAITGDHANLRVYIGTNGGIYVGD 780
 Db 706 AEIGGVGRGIFRSTDAATWTRVNDDAHQWMTAAITGDPVYGRVYVATNGKVIYGD- 764
 Qy 781 GGAPEGSPSPSVSPSASPSTSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840
 Db 765 -----TSDTGGGTDPEGPDPPT----- 782
 Qy 841 SASPSPSSSPSPSSSP 900
 Db 783 -----TGACEVTYVTNQMWPBGQF--ADVRLTNTGT 811
 Qy 901 S-----SVLS-----TVYRVYFTR--DGSSSLVYNCDM-----AATGCGNIRA--- 939
 Db 812 SAMNGWSTLWSPSPGQGEVTRMNAEHTQAQTSVTAARVGNAGVAPGASVGFPTGSRSG 871
 Qy 940 -----SFGSVNPAATPTA 951
 Db 872 TTAEPGEFVAVAGRACPTA 889

RESULT 2

D97013
 probably secreted sialidase, several Asp-boxes and dockerin domain [imported] - Clostrid
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
 C:Accession: D97013
 R: Nolling, J.; Breton, G.; Omlachenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J.; Bacteriol. 183, 4823-4838, 2001
 J: Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
 A:Reference number: A96900; PMID:21359325; PMID:21359325
 A:Accession: D97013
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-839 <KUR>
 A:Cross-references: UNIPROT:Q97JKO; GB:AE001437; PTDN:AAK78895.1; PID:g15023820; GSPDB:C
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC0919

Query Match 39.7%; Score 2037; DB 2; Length 839;
 Best Local Similarity 49.5%; Pred. No. 1.2e-83;
 Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;
 Qy 13 SRRVLSLAA-----TASPAVAALGVLPIATSPAHAAAT--TOPYTSNVAIGGGGVD 66
 Db 2 NKRIIVSMVAGLSITFTTG-----VTHISAANKKAAASVQSGYKMDAKITAGGIYP 54
 Qy 67 GIVNFGAPGILYVRDIDIGYRMDANGRMIPLLDVGWNNWNGYGVASIAADPINTK 126
 Db 55 AVIFNTEKOLIVARTMGAYRMDKANKNMIPITD--GFSDWMTLGCESIAIDPIDNR 112
 Qy 127 WAAVGMVITNSWPNDCAILRSSQAGTWTQITPPLFKLGGNMPGRGKERLAVDPNNNI 186
 Db 127 WAAVGMVITNSWPNDCAILRSSQAGTWTQITPPLFKLGGNMPGRGKERLAVDPNNNI 186

Db 113 VYIAAGLYTNDMDENAVILSSODKNTWKRYQLPFKVGNGMPPGRNGERLQIDPNDKI 172
 Qy 187 LYFGAPSGKGLWRSTDSGATMSQMTNPPDVGTYLANPTDTGYOSDIGVWVAFDKSSS 246
 Db 173 LYLARSGNGLMKSEDIQGTWISKVDNFPDIDYQDPQNE--YIADKXGVWWEFTDPSTG 230
 Qy 247 SLQASRTIFGVAD--PNNPVFMSRDGATQAVGAPGTGFIPIKGVDPVNVHLYIATS 305
 Db 231 TKSGPQTQMYGAADKTKGNNTYYTNDGKMTSAKQKGLPLPHNGIL--ABDGMLYISVS 289
 Qy 306 NTGPEYDSSGSDVWKFSVTSCTWTRISFPVSTDIANDFYSGSLTIDRQHPNTIYVAIQ 365
 Db 290 NTGPEYDSSGSDVWKFSVTSCTWTRISFPVSTDIANDFYSGSLTIDRQHPNTIYVAIQ 365
 Qy 366 SWWDPTLIFRSTDDGATWTRIMDTSYNRLRYLDOIASBPWLTFG--VQNPVPVSPPL 424
 Db 347 RWWDEELIYRSTDAKTKWPIIMDNGVYRNRLGYNLDVSAQPMWDMGTGVTPEPLVYL 406
 Qy 425 GMDENAMIDPFNSDRMLYGTATLYATNDLTKMDSGQIHIAPWVKGLETAVNDLISP 484
 Db 407 GMMGDLIDPFNSDRMLYGTATLYGTDLTMDKGNVDSIYVANGIEBCAVADVVP 466
 Qy 485 PSAPLISALGDLGGFTHADVTAVPSTIFTSPTVFTTGTSDVYAEINPSIIYRAGSPDPS 544
 Db 467 TKGAQLISAVGDDCGFYHDDITKVPSSKMTTPNFSATTSIDYAESVPMFVVRGVNDTSK 526
 Qy 545 QPNDRHVAFSDGGKNNWQ--GSEPGVTTGTVAAASADGSRFVWAPGDQGPVVYVAVGRG 603
 Db 527 NQDKKDCGISYDGGKNNWSP--SGVYKAGTVAAAGAKTIWSP--EGANAAASTDNG 585
 Qy 604 NSWAASQGVANAOIRSDRVNPKTFYALSGNFTYRSTDDGVTPEQVAA--GLPSSGAVWVF 662
 Db 586 NKMTPTCSGLPGAKVRSRVPKPFYGLNGKFIYSTAGATFTTOSQGLPTKRG--GI- 643
 Qy 663 FHAVPKGEGDLMLA--ASSGLYHSTNGSSWSAITGVSSAVNVGFKGAPGSSYPAVFVVG 721
 Db 644 FKTYIGHGDIWAGKRGDLHSTDSGATFTKSVSDASDPTVGLKSKTKDDGYPAIYMDA 703
 Qy 722 TIGGVTAAYRSDDCCTTWVTLINDDOHYGNMGQAITGDHANLRVYIGTNGGIYVGD 781
 Db 704 TIDGAGCIFRSDDEGATWTRINDDAHQYGSPPYCTIDPKNKXGVFTNGRGIYVGDID 763
 Qy 782 GAGSGSPSPSVS 793
 Db 764 GS-QPTPSPSVT 774

RESULT 3

T00349
 Avicelase III - Aspergillus aculeatus
 C:Species: Aspergillus aculeatus
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
 C:Accession: T00349
 R: Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.
 submitted to the EMBL Data Library, June 1998
 A:Description: Avicelase III from Aspergillus aculeatus.
 A:Reference number: Z14141
 A:Accession: T00349
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-856 <ARA>
 A:Cross-references: UNIPROT:O74170; EMBL:AB015511; NID:d1199887; PID:d1029971
 C:Genetics:
 A:Gene: avIII1
 F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 33.9%; Score 1740; DB 2; Length 856;
 Best Local Similarity 43.5%; Pred. No. 2.1e-70;
 Matches 366; Conservative 139; Mismatches 297; Indels 40; Gaps 18;
 Qy 30 AALGVLPIATITASPAHAATTOPYTSNVAI--GGGFFVGGIVFNEGAPGILYVRTDIGMY 88
 Db 4 SSLALICALLGLKADDAASQAYTKNVYTGSGGFTPTQIVNPSAKGVAVARTDIGAV 63

QY 89 RWDANGREWILLDMVG---KNNWYGVVSIADPINTNKVMAAVGMYTNSWDPNDAI 145
DB 64 RLN-SDDTWTPLMDVWGNWDTWHDW---GIDALATDPVTDRIYVAVGYTINWEDNVGSI 119
QY 146 LRSSDOGATWQITPLPFKLGCMRPGRMGERLAVDPNNNDNLIYFGAPSGKGLWRSTDGA 205
DB 120 LRSTDGDTWTETKLPFKVGGNMPGRMGERLAVDPNNKNSLIYFGARSGHGLWKSTDYGA 179
QY 206 TWSQMTNPPDVGTYIANPTDTTGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNP 265
DB 180 TWSNVTSTFTGTTFQDSSST--YTSDDPVIWVTFDSTSGSSSATPRIIFGVADAGKS 237
QY 266 VFWSDGATQWAVGAPT-GFIPHKGVDPVNVHLYIATNTGPGYDSSGDVWKFVST 324
DB 238 VFKSDAGATWAVSGEPQYGLPHKGVLSPEEKLTYISYANGAGPYDGTNGTVHKYNI 297
QY 325 SCTWTRISVPFSTDTANDYFGVSLTIDROHPNTIIVATQISWPDITIIFRSTDGATWT 384
DB 298 SGVWTDIISF---TSLASTYGYGGLSVLDQVPGTLMAALNCWNPDELIFRSTDGATWS 354
QY 385 RIWDTSYPNRSLRLVLDISAEPLTIFGVQPNP-FVPSPKLGWMDMAIDPFNSDRMLY 443
DB 355 PIWEKNGYPSINYYSYDISNAPWQDITSTDPFV---RVGMVYEAALAIIDPFDSNHWLY 411
QY 444 GTGATLYATNDLTWKDSCGQIHIAPMVKGLBERTAVNDLISPPSGAPLISALGDLGFTHA 503
DB 412 GTGLTVYGGHDLTNWDSKHNVTVKSLAVGIEBMAVLGLITPPGGPALLSAVGDGDFYHS 471
QY 504 DTVATVPSTIFSPVPTTGTSDVYAEINPSIIVRAGSFDPSSOPNDRHVAFTDGGKNWFQ 563
DB 472 DLDAAPNOAYHTPTGTGTNGIDYAGNKPNSIIVRSGASD--DYPT---LALSNFGSTWYA 526
QY 564 GSEPGVTTGGTVAAASDRSFWAPDGPQVAVVAVFGNSWAAASQGVPAANAQIRSDRV 623
DB 527 DYAASTSTGTGAVALSADGDTVLLMSSTSGALVSKSQG---TLTAVSSLPSCGAVIASDKS 583
QY 624 NPKTFYALNSGTYFRSTDGGVTFQVPAAGLPSGSAVGVMFHAPVKGKGLWLAASGLVH 683
DB 584 DNTVYGGSGAGAIYVSKNTATSTFTKTVS-LGSSTTVNAI-RAHPSIAGDVMASTDKGLWH 641
QY 684 STNGSSWSAI-TGVSSAVNVGFGKAPGSSYPVAVVGVGTIGVTVGAYRSDCGITWLI 742
DB 642 STDYGSTFTIGSGVGTAGMSFGKASGTGSVWYIGFTIDCAAGLPKSEDAGTNQVVI 701
QY 743 NDDQHQYGNWGA--ITGDHANLRVYIGTNGRGIYVGDIGAPSG-----SPSPSV 792
DB 702 SDASHGFGS-GSANVNGDLQTYGRVGRHERPGLLRQSQREPAGRHGDGDDTTYSKT 760
QY 793 SPSSASPSLSPSSPSPPSSPSPPSSPSPPSSPSPPSPSPSPSPSPSPSPSPSPSPSP 852
DB 761 STTVSTLLKTTSSASTSSSTTVKTTSSSTTSKASSTTTTSTTTTSSGTTATA 820
QY 853 SS 854
DB 821 SA 822

RESULT 4

F72393
hypothetical protein TM0305 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: F72393
R: Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; PMID:99287316; PMID:10360571
A:Accession: F72393
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-707 <ARN>

A:Cross-references: UNIPROT:Q9WYEL; GB:AE001712; GB:AE000512; NID:G4980799; PIDN:AAD353

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0305

Query Match 22.6%; Score 1161; DB 2; Length 707;
Best Local Similarity 35.2%; Pred. No. 9.9e-45;
Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;
QY 52 YTWSNVATGGGFGVGIIVNEGAPGLIYRTDIGMYRWDPAANGRWIPLLDWGWNNWGY 111
DB 21 FEWKSVEINGGFGVPIIIFHPASGLLYARTDVGGLYRWDDEETKRWKQLDFLRDQSDY 80
QY 112 NGVVSIAADPINTNKVMAAVGMYTNSWDPNDAILRSSDOGATWQITPLP---FKLGNM 168
DB 81 MGVLSVALDPSPKRIYAMTKYTQDW-AGYGAILISEDYGETWIVNLKYGIKVGNE 139
QY 169 PGRGMGERLAVDPNNNDNLIYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTT 228
DB 140 DGRNAGERLOVDNPFSSVLFMGT-TKYGLWKSEDFGKNWKKVDSFPST----- 186
QY 229 YQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPVPFWSRGGATWQAVGAPTGP 288
DB 187 -----SVTFVLDFEKSGEKSPTRIFVGCSEPKG-IFVTEDGGTTWNLPLNDLIP 239
QY 289 HKGVDPVNVHLYIATNTGPGYDSSGDVWKFVSTGTWTRISVPFSTDTANDYFGVSG 348
DB 240 LRGLKH--DGILYVTLNALGPNAGATGAMKXVIADQKWDYVTPMKGD-----FGYCG 291
QY 349 LTIDRQHPNTIIVATQISWPDITIIFRSTDGATWTRIMDWTSYPNRSLRLVLDISAE 408
DB 292 IDVQE---NVIVSTLDRWYPHDEIFISLNGGETWRLLEKANF-----DINKAP 339
QY 409 LTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWKDSCGQIHIAP 468
DB 340 IK---DLNP-----HWISD-VKIDPFDMNRAIFTTGYGVWVYELKKSPEG---MGK 384
QY 469 MVK-----GLETAVNDLISPPSGAPLISALGDLGFTHADVTAVPSTIITSVPFTGT 522
DB 385 PVKWIPENRGLEETVVLQVLPPIGERPILLSAIDWGGFRHESLDTPTSSMY-KPLKWTSL 443
QY 523 SDVYAEINPSIIVRAGSFDPSSOPNDRHVAFTDGGKNWFQ-GSEPGVTTGG--TVAA 579
DB 444 GIATFAYQNSKFAVHTY---TYP---FLSYSDGGINWREIETVPEGITDGGRLSLAV 497
QY 580 ADGSRFVWAPDGPQVAVVAVFGNSWAAASQGVPA---NAQIRSDRVNPKTFYAL--SNG 634
DB 498 NDKTILWSPAN--HEVIVSSDKSKWKKAISVPVPEFNPFPASDPVNPSPFYIFDWKNG 555
QY 635 TFYRSTDDGVTQFVAAAGLPSGSAVGVMFHA-----VPGKGDLMALAS-SGLYHSTNGS 689
DB 556 DFLISKDGGKSPMK-GAKLPSPDNWVSLYSFPVLAADREGDIWLALQWGLYRSKDG 614
QY 690 SWSAITGVSSAVNVGKSPAGSSYPVAVVGVGTIGVTVGAYRSDCGTTWVLINDDHOY 749
DB 615 TFERLGNVDIAYVIGFAPKEGTDYPAIYLVNGMVNGVYIGFIMSTDEGATWIRINNDK 674
QY 750 GNMGQAITGDHANLRVYIGTNGRGIYVGD 780
DB 675 G-WIHYMIGDMNEFGRIFLGTGEGRIIVGEV 704

RESULT 5

T31085
xylanase - Caldicellulosiruptor sp.
C:Species: Caldicellulosiruptor sp.
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004
C:Accession: T31085
R: Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.
submitted to the EMBL Data Library, December 1997
A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.
A:Reference number: Z20972

Db	253	KWAKCWDVHNGAAL--LLAKITGKIYKQIIIESHLDDWITGNGERIKYTPKGLAWLD	309
Qy	564	--GSEPGGVTTGGTVAASAD-----GSRFVWAPGDPGPVVAVVAVGFG--	603
Db	310	QWGLSLRYATTTAFLAVFVSDWVGCPSSTKKEIYKFGESQIDYALGSAGRS--FVVGFGTN	367
Qy	604	-----NSWAASQGVPA-----NAQIRSD--	621
Db	368	PPKRPKPHRTAHSSWADSQISPIYHRTLYGALVGGPGSDSDYDDISNVNNEVACDYNA	427
Qy	622	-----RVNP-----KTFYALSNCTFYRSTGGVTFQPVAAAGLPSFGAVGVMPH	664
Db	428	GFVGAALKMYQLYGNGNIPDFKAIETPTNDEFF-----VEAGINASGTNFIETK	476
Qy	665	AVPGKEGDIWLAASS-----GLYHSTNGSGSSWALITGVSSAVNVGFGKAP-----	710
Db	477	AIWNQSG-WPAKATDKLFRYFVDLSLSEIKAGYSPNQLT-LSTNTNQAKVSGPVVMDA	534
Qy	711	--GSSYPVAVVGTI--GGVTCAYRSDDCGTTWVLINDDOHQYGNWQALITGDHANLREV	766
Db	535	SKNIYYILVDFGTGLIYPGGQDKYKE-----VQFRIAAPQNVQNDNSYDFQDIKV	588
Qy	767	YIGT-----NGRGIVYGDIGGAPSGSPSVSPASPSLSPPSPSPSPSPSP	816
Db	589	SGSVVVKTKIPILYDGVKVGWDGPGTSGATPTPA-----TATPTPTPTPTPTPTPT	642
Qy	817	SSPSPPSP	875
Db	643	TS--TATPT	690
Qy	876	QYKNDNSAPGDNQIKGLQVNTGSSVDLSVTYVYVTRDGGSTLVNCDWAAGCG	935
Db	691	LYANKETNTTNTIRPWLKVNTGSSIDLRSVIRYVYVTRDGGSTLVNCDWAAGCG	749
Qy	936	NIRASPGSVNTPATPTADTYLQ	956
Db	750	NVTFKFLSSSVSGADYLYE	770
RESULT 7			
Tl7120			
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum			
C;Species: Caldocellum saccharolyticum			
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C;Accession: Tl7120; A43745			
R;Te'o V.S.; Saul, D.J.; Bergquist, P.L.			
Appl. Microbiol. Biotechnol. 43, 291-296, 1995			
A;Title: Cella, another gene coding for a multidomain cellulase from the extreme thermoph			
A;Reference number: Z18698; MUID:95336703; PMID:7612247			
A;Accession: Tl7120			
A;Status: preliminary; translated from GB/EMBL/DBJ			
A;Molecule type: DNA			
A;Residues: 1-1742 <TEO>			
A;Cross-references: UNIPROT:P22534; EMBL:L32742; MID:g537499; PID:g537500; PIDN:AAA91086			
R;Luethi, E.; Bhana Jagmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.			
Appl. Environ. Microbiol. 57, 694-700, 1991			
A;Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding			
A;Reference number: A43745; MUID:91247819; PMID:2039230			
A;Accession: A43745			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1516-1544, 'A', 1546-1742 <LUE>			
A;Cross-references: EMBL:M36063; MID:g144292; PIDN:AAA72860.1; PID:g144293			
C;Genetics:			
A;Gene: Cella			
C;Keywords: glycosidase; hydrolase			
Query Match 7.7%; Score 397.5; DB 2; Length 1742;			
Best Local Similarity 21.7%; Pred. No. 2.7e-10;			
Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;			
Qy	262	PNNVPWRSRGGATWQA---VEGAPTGFPHKGVDFPVNHLVY-IATSNTPGFGYDGGSG	316

Db	48	PN-----WVRN-----NWRGDSALKDQDNGLDLTCGFWFADGDKVFNLPMSYTGTMLSWAA	100
Qy	317	DWVKFSVTGTRISPVSPSTDTANDYFCVYSGLTIDROHPNTIMVATQI-----SWW-	368
Db	101	EYKDAFVKSQGLEHI--LNQIEWNDYF-----VKCHPSKVVYVQVGDGKDHAWWG	151
Qy	369	-----PDTIIFRSTDDGATWTRIMDWTSPNRSRLRYLDISAEPMLTIFGVQVQNPVP	419
Db	152	PAEVQMWERPSKVTQSSPGSA-----VVAETAASLAASIVLKDRN	193
Qy	420	PSPKGLWMDMAIDPFNSDRMLYGTGATLYATNDLTWKDSCGQIHIAPMVKGLEETAVN	479
Db	194	PTKAATYLGHA-----KOLYBFAEVTKSDSG-----YTAAN	224
Qy	480	DLISPPSG-----APLISALDGLGFTHADVTAVPSTIFTPSVFTTGTSTVDY---	526
Db	225	GYTNSWSGFDLSMAVWLYLATNDSTYLTKAE-----SYQNWPKISGNSNIIDYKWAH	279
Qy	527	-----AELNPSIIVRAGSPDPSPQPNDRHVAFTDG-----GKNW-----	561
Db	280	CWDDVHNGAAL---LLAKITDKYQIIIESHLDDWITGNGERIKYTPKGLAWLDQWGS	336
Qy	562	-----FQSGEPGGVTTGGTVAASADG--SREVNAPGDPGPVVVAVGFG-----	603
Db	337	LRYATTTAFLAVFVSDWVGCPSSTKKEIYKFGESQIDYALGSTGRS--FVVGFGTNPPKR	394
Qy	604	-----NSWAASQGVPA-----NAQIRSD-----	621
Db	395	PHHRTAHSSWADSQISPIYHRTLYGALVGGPGSDSDYDDISNVNNEVACDYNA	454
Qy	622	-----RVNP-----KTFYALSNCTFYRSTGGVTFQPVAAAGLPSFGAVGVMPH	668
Db	455	ALAKMYLLYGGNPIPDFKAIETPTNDEFF-----VEAGINASGTNFIETKAI--	501
Qy	669	KEGDLMLAAASGLYHSTNGSGSSWA-----ITGVSSAVNVGFGKSGAPSSYPAVF	718
Db	502	-----VNNQSGWPARATNKLKFRYFVDLSLSEIKAGYS-----ENQL	537
Qy	719	VVGTI-----GGVTGAYRSDCGTTWVLIND-----DQHQYGNWQALITGDHANLRR	765
Db	538	TLSTNTNQAKVSGPVVWDSSRNIIYILVDFGTGLIYPGGQDKYKKEVFRIAPQNVQ-	596
Qy	766	VYIGTNGRIVYDIDGAPSGS-----PSPSVSPASPSLSPPSPSPSPS	807
Db	597	---WDNSNDYSFQDIKGVSSGVVTKYIPLYDEDIKVMGEEPGTS-----GVSPTPTAS	648
Qy	808	SSPSPPSP	867
Db	649	VTPPTPTPTPA-----TPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT	700
Qy	868	P-VSGGVKVOYKNDNSAPGDNQIKGLQVNTGSSVDLSVTYVYVTRDGGSTLVN	926
Db	701	PATSGQIKVLYANKETNTTNTIRPWLKVNSGSSIDLRSVIRYVYVTRDGGSTLVN	760
Qy	927	CWMAALCGGNIRASFGSVNTPATPTADTYLQ	956
Db	761	-DWAQIGASNVTPFKVKLSSSVSGADYLYE	789
RESULT 8			
S50755			
hypotheical protein VSP-3 - Chlamydomonas reinhardtii			
C;Species: Chlamydomonas reinhardtii			
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004			
C;Accession: S50755			
R;Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.; Haring			
Plant Mol. Biol. 26, 947-960, 1994			
A;Title: Domain conservation in several volvoclean cell wall proteins.			
A;Reference number: S50754; MUID:95093034; PMID:8000007			
A;Accession: S50755			
A;Status: preliminary			
A;Molecule type: mRNA			

T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T34434
R:Geisler, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans coamid K06A9.
A:Reference number: 221525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: UNIPROT:Q81FK6; EMBL:U08846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 5.7%; Score 290.5; DB 2; Length 2232;
Best Local Similarity 21.8%; Pred. NO. 2.1e-05;
Matches 239; Conservative 114; Mismatches 428; Indels 315; Gaps 45;

QY 18 SLLAATASFAVAALGLVPIA---ITASPAHAATTQPYTWGNVAI---GGGPFVDGIVENE 72
DB 235 SLGTSSSLPSSISTALPIASSASSSPSAASSTPVVLSSSTIQSSSGTFPPSSVASSP 294

QY 73 GAPGILYVRTDIGMYRMDAANGRIPLLDVWGNWNGYVVSIAAD-----PINTNK 126
DB 295 STVG-----STSG-----AASSSYATVSTIAGTGTITFPVPGSS 330

QY 127 VMAAVGYMTNSWDPNDAILR---SSDQAGTQITP-LPFLKLGNNMP---GRGMGERLAVD 180
DB 331 ---STIGSTPSASSSSSGTMTISGSGTGVTVVPGSSSTPASSTPIASSSSPGSTVIVA 388

QY 181 PNNONILYFGAPSGKGL---WRSDSGATWSQMTNPPDVGVIANPTDTTGYQSDIQGV 237
DB 389 PGSSSTYGSSTPSASSSSSGTMTNSGSGTGVTVVAPVSSSTFGSGSTPIASSSSSGSTV 448

QY 238 WVAFDK-----SSSSLQAS-----KTFVGVADPNPVFWRDGGATWQA 278
DB 449 VSGSSSTYGSSTPSASSSSAGTASTISGSGSTATVIVPGSS-----SSVGSSTQSA 500

QY 279 VPGAPTGFIPHKGVDPVNVHLYIATNTGGPYDGGSDGVKPFVTSGTWTRISFVPS 338
DB 501 SPSSPG-----TWSTVSGP-TGSGTVTVPGSSTSPA-PSSSNPSSS 540

QY 339 TANDYFGYGLTIDRQHPNTIMVATQISHWPDITIFRSTDDGATWT---RIWDWTSYPNR 395
DB 541 PAS-----TGSTTITSGSSSIIVST-----VSGSTVSGTGTSTQSTLASSTATPGS 586

QY 396 SLRYVLDISAE-----PWLTFGVQPNP---PVSPKLGWMDAMADPPNS 438
DB 587 SSTVPSSSSPQSSQSPAPNTGSTTSPQSTSSQSPSPMNPSSSTPTGSSQSTTITPEGSTA 646

QY 439 DRMLYGTGATLYATNDLKWDSGGQIHIAPMVKGLEETAVNDLISPPPGAPLISALGDL- 497
DB 647 SSPTGSGTSTSVATEVTSQST-----VPSGSSLGQTQSTNSSPSPSSLSPTSGMSTLT 700

QY 498 -----GGFTHADVAVPS-----TIFTSVPVFTTGVSDVYAEINPSSIIVRAGS 539
DB 701 SEPPSPSTQSSGAQSTLTTPSPNPQSTSSLSBSSTSGATTSSGSAGTTMTSPSQSSSVGS 760

QY 540 FDPSSQPNDRHVAFSTDGKWFQGS--EPG-GVTTGGTVAASADGSRFVWAPG-----590
DB 761 SQGSTSP-----AASSTGEMTSGSTQTPGSSVSTSAAILTSTQOSVSTNSPGSTVTRAP 815

QY 591 -----DPGQPVVY-----AVGFGNSWAASQGVAPNAQIRSDRVNPKTFYALSNGTFFYRS 639
DB 816 STVSGSTSSGSGTTVTGSTFASSTGSSVASSPAPTSQ-----NPNPSTSSGSSMITQS 869

QY 640 TDGGVTFQFVAAGLPSSGAVGYMFPHAVPGKEGDLMLAASSGLYHSTNGGSSWSA-----IT 695

Search completed: October 5, 2005, 07:57:01
Job time : 32.555 secs

Db 870 PYPQSTSTSPVESSTTPS-----PGSPGTTLTSTSPSPSQSTTIGSTQGSTSPGIS 919
QY 696 GVSSAVNVGFGKSAFGSSYPVAVVGTIGGVTCAYRSDDCGTTWLVINDDDHQHYGNWQA 755
Db 920 TTSEMTSQGSTQTPGSGSTGTQSTVSDST-----951
QY 756 ITGDHANLRRVYVGTNGRIGVYDIDGAPSGSPSPSPSASPSPSPSPS-----PSSSP 810
Db 952 -----SSGSTVTVGSTEK---SSSPISTQNTNPSTSSGSSMSSTQTPQSSQ 995
QY 811 SPSP---SPSSSPSSSPSP-----SPSPSPSP-----SPSPSPSPSS 848
Db 996 STSPVESSTSGATSSSGSPGTTLTISISPSPSSTIGSSQGSTSPVVTISQGSTETPGS 1055
QY 849 SPSPSSSP-----SSSPSPSPSSSPSPSGVGVKVKYKND 881
Db 1056 TGSTVTKPSTVSGSASSGSTATWGTASTGSGSTSPNPQSTSPSTSGA-----T 1108
QY 882 SAPGDNQIKPGLQVNVNTGSSSVDLSTVTVRYWFTRGSGSSTLVYNC--DWAALCCGNIRA 939
Db 1109 SPFGSS---GTTLTISISPSQSGSTIG-----SSQGSTSPVVTISGDMTSQGSTOIPG 1159
QY 940 SFGS--VNPATPTADT 953
Db 1160 STGSTVTPQSTGSGST 1175

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:44:57 ; Search time 118.987 Seconds
(without alignments)
4118.611 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLTWRSRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2479.5	48.3	882	Q82M04	streptomyc
2	2436	47.4	890	Q86727	streptomyc
3	2420.5	47.1	996	Q9AQH0	caldicellul
4	2037	39.7	839	Q97KK0	clostridium
5	2007	39.1	842	Q70DK5	clostridium
6	1831.5	35.7	857	Q8TFP1	aspergillus
7	1784	34.7	806	Q9P4T8	agaricus bi
8	1740	33.9	856	Q74170	aspergillus
9	1625.5	31.7	838	Q729M8	trichoderma
10	1587.5	30.9	861	Q7S1W6	neurospora
11	1442	28.1	739	Q82K30	streptomyc
12	1173.5	22.9	751	Q8FLM5	xanthomonas
13	1161	22.6	707	Q9WYE1	thermocoga
14	1151.5	22.4	751	Q8P9U5	xanthomonas
15	1125	21.9	812	Q8J0D2	geotrichum
16	993	19.3	776	Q764N8	geotrichum
17	431.5	8.4	1779	Q52374	caldicellul
18	409	8.0	196	Q8J1H7	agaricus bi
19	409	8.0	1770	Q9X3P5	caldicellul
20	402.5	7.8	921	Q918L8	caldicellul
21	398.5	7.8	1711	Q96311	anaerocellu
22	397.5	7.7	1742	Q22534	caldocellum
23	387	7.5	2014	Q7U3X4	synecococc
24	385	7.5	1664	Q7X2U2	uncultured
25	372.5	7.3	473	Q39620	chlamydomon
26	360	7.0	577	Q6FSJ1	candida gla
27	354.5	6.9	790	Q6FX25	candida gla
28	350.5	6.8	997	Q924I1	bacillus sp
29	346.5	6.7	901	Q44562	actinomyces
30	340.5	6.6	1331	1 MAMB	caldocellum
31	336.5	6.6	741	Q82QF2	streptomyc

32	335	6.5	911	2	Q72WN1	desulfovibr
33	334	6.5	1751	2	Q9AQG4	caldicellul
34	325.5	6.3	930	2	Q9RFX5	caldicellul
35	325	6.3	1000	2	O24820	thermophili
36	320.5	6.2	991	2	Q94C44	chlamydomon
37	318.5	6.2	875	2	Q9F2B0	thiobacillu
38	316	6.2	1091	2	Q8KKF7	paenibacill
39	308.5	6.0	611	2	P74375	synecocyst
40	301.5	5.9	1749	2	P73032	synecocyst
41	298.5	5.8	1154	2	Q7U5X7	synecococc
42	297	5.8	714	2	Q7U5X6	synecococc
43	297	5.8	1915	2	Q8RPL0	acetivibrio
44	294	5.7	474	2	Q86154	dictyosteli
45	293	5.7	279	1	Y091_NPVP	orgyia pseu

ALIGNMENTS

RESULT 1

ID	Q82M04	PRELIMINARY;	PRT;	882	AA.
AC	Q82M04;				
DT	01-JUN-2003 (Tremblrel. 24, Created)				
DT	01-JUN-2003 (Tremblrel. 24, Last sequence update)				
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)				
DE	Putative endo-1,4-beta-glucanase.				
GN	Name=celA3; OrderedLocustNames=SAV1856;				
OS	Streptomyces avermitilis.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OX	Streptomycineae; Streptomycetaceae; Streptomyces.				
NCBI	NCBI_TaxID=33903;				
[1]					
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MA-4680;				
RX	MEDLINE=22608306; PubMed=12692552;				
RA	Ikedda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,				
RA	Sakaki Y., Hattori M., Omura S.;				
RT	"Complete genome sequence and comparative analysis of the industrial				
RL	microorganism Streptomyces avermitilis.";				
Nat.	Nat. Biotechnol. 21:526-531(2003).				
[2]					
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;				
RX	MEDLINE=21477403; PubMed=1157948; DOI=10.1073/pnas.211433198;				
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,				
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,				
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;				
RT	"Genome sequence of an industrial microorganism Streptomyces				
RL	avermitilis; deducing the ability of producing secondary				
metabolites.";					
Proc.	Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).				
EMBL	AF005028; BAC69567.1; -.				
HSSP	P07986; 1EXG.				
GO	GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.				
GO	GO:0005575; P:carbohydrate metabolism; IEA.				
InterPro	InterPro: IPR001919; Bac celose-bind.				
InterPro	InterPro: IPR008965; Cellul bind.				
InterPro	InterPro: IPR002860; Glyco hydro_BNR.				
InterPro	InterPro: IPR011040; Sialidase.				
Pfam	Pfam: PF02012; BNR. 10.				
Pfam	Pfam: PF00553; CBM_2; 1.				
SMART	SMART: SM00637; CBD_II; 1.				
KW	Complete proteome.				
SEQUENCE	882 AA; 92152 MW; A3C8E753CE69F13D CRC64;				

Query Match 48.3%; Score 2479.5; DB Z; Length 882;
Best Local Similarity 50.1%; Pred. No. 1.5e-96;
Matches 476; Conservative 129; Mismatches 230; Indels 115; Gaps 15;
QY 11 MRSRRRLVSLAATAATPAFAAAL--GVLPPIAITASPAHAATTQPYTWSNVAIGGGGFVDGI 68
Db 1 MRRTRIF-----TAVLALAGLPAGTTPALAAAPTATIAADTYTSWKNARVDGGGFVGI 55

QY	69	VFNAGPGLYVRDITGGMYRWDAANGRWIPLLDWGWNNGYNGVSVIAADPINTNKVW	128
DB	56	VFNRSKNIAYARTDIGAYRAEESKWTPLLDVSGWSDGHTGTVVSLASDSVDPNKVY	115
QY	129	AAVGMYTNSWDPNGAIIILRSSDQAGATWQITPLPFKLGGMGRGMRGERLAVDPNNNDILY	188
DB	116	AAVGYTINSWDPNGAVLRSDDRGASQKTDLPFLKLGGMGRGMRGERLAVDPNNNSVLY	175
QY	189	FGAPSGKLGWSTDSGATWSQWNPDPVGTIANPTDITGVQSDIQGVVWVAFDKSSSL	248
DB	176	LGAPSGKLGWSTDSGASWSQVTPFNVGTYYQDADITSGVASDNQGIWVTFDESTGSP	235
QY	249	GOASKTIFVGVADPNPNVFWSRDGGATWQAVPGAPTGFPHKGVDPDPVNNHLYIATSNIG	308
DB	236	GSSTRTVYGVADKDNSYRSTDAAGTWSRLAGQPTGHLAHKGVLDAAAGCLYLAYSQK	295
QY	309	GPYDSSGDVWKFVSTGTRISVPSTDTANDYFGYSGLTIDRQHPNTIMVATQISW	368
DB	296	GPYDGGKQLWRYTKTGTWINSVPAEADT---YGFSGLTVDQRHPTGMATAYSSW	352
QY	369	PDTIIFRSTDGGATWTRTMDWTSYFNRSIRYVLDISAEPLTFFGVOPNPVPSPKLGWMD	428
DB	353	PDTQLFRSTDSGGTWTAKWDYTSYPSRNRFTMDVSSSGPWLTMGANPAPPEQTPLKGMWT	412
QY	429	EMALDPPNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLISPPS-G	487
DB	413	ESLEIDPFDSDARMVYGTATVGTUNLTWDSGSOFTIKPMARGLEETAVNDLASPPSG	472
QY	488	APLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDVYAEINPSSIIVRAGSFDPSQPN	547
DB	473	AQLFSALGDIIGFRHTDITVPSLMTYSPNFTTSLDYAETDPTGVVRVGNLD--SGP-	529
QY	548	DRHVAFTDGGKWNWQGSPEGVTGCTVAAASDGRFVWAPCDPGQVPVYAVGNSWA	607
DB	530	--HVAFTDNGANWPAFADPSGSGGTVAASDGRFVWSPAGTG--VQYTTGFTGWS	585
QY	608	ASQGVPAANAQIRSDRVNPKTFYALNGTFFYRSTDCGVTPOFVAA--GLPSSGAVGVWFAV	666
DB	586	ASAGLPAGAVESDRVDPKTFYFGSGRFPYVSSDGGATFASATGLPSGDS--VRFKAL	643
QY	667	PKRGDMLAASS-----GLXSTNGSGSSAITGVSSAVNVGFGKSAPGSSYPVAVVVG	721
DB	644	PGTKGDIWLAGSGADGAYGLWHSTDGGAAFTKLTAVDQADTIGFKAATGASYQTLTSA	703
QY	722	TIGGVTGAYRSDCCTWVLINDHOQVGNWQATIGDHANLRVYIGTNGRGIVYGDIG	781
DB	704	KITGGVGRGIFRSTDKGASWTRVNDDAHQWGTGAATIGDPRVYGRVYVSTNGRGIVYGDTA	763
QY	782	GAPSGSPSPSVSPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS	841
DB	764	GSSDGG-----	769
QY	842	ASPSPPSP	901
DB	770	-----GTEPAPT-----GACTVYVIRITNQSSGGFQ--ADVQLANTGST	805
QY	902	SVDLSTVTVRYWFTRDGSSSTLVNCNDWAAICG-----NIRASFGS	943
DB	806	AMDGWSLG---WSFGDGEVETQLMNASYAQAQSGVTRANLAWNGRVAAGS	852
RESULT 2			
ID	OB86727		
AC	OB86727	PRELIMINARY; PRT; 890 AA.	
DT	01-NOV-1998 (TreeBLrel. 08, Created)		
DT	01-NOV-1998 (TreeBLrel. 08, Last sequence update)		
DT	01-MAR-2004 (TreeBLrel. 26, Last annotation update)		
DE	Putative secreted cellulase.		
GN	ORFNames=SC5C7.30c;		
OS	Streptomyces coelicolor.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		

Qy	547	NDRHAFSTDGKKNWFQSGSEGGVTTGTTVAASADGSRFVWAPDGPQPVVYVGFQNSW	606
Db	533	---HIAFSTDNGANWFQGTDFSGVGGCTVAAGADGSRFVWSP--EGAGVQYTTGFGTSW	587
Qy	607	AASQGVANAIQRDRVNPNTFYALSNGTFRSTGDTGVTPOVAA-GLPSSGAVGVMFHA	665
Db	588	QASTGLPAGAIVESDRVNPATFYGKSRFVYSTDGGATFTASAATGLPAGD--GVREKA	645
Qy	666	VPKEGDLAASS-----GLYHSTNGSSNSAITGVSSAVNVGFGKSAPGSSYPAVFVW	720
Db	646	LPGEGDWMLAGGAADGYPGLWHSTDDGGTFTTRUPGVDAADTVGFGKAAPGASVQTLFTS	705
Qy	721	GTIGCVTCAYRSDDCGTTWVLINDDHOQYGNWGOAITGDHANLRVYTGTCRGIVYGD1	780
Db	706	AEIGVRCIFRSTAGATWTRVNDDAHQWGTGAITGDPVYGVYVATNGRGIYGD-	764
Qy	781	CGAPGSP	840
Db	765	-----TSDTGGTDPGPGDPTP-----	782
Qy	841	SASPSPPSP	900
Db	783	-----TGACEVYTVTNQWPGGQFQ--ADVRLTNTGT	811
Qy	901	S-----SVDLN-----TVTVRYWFTF--DGGSTLVYNCW-----AAIGCGNIRA---	939
Db	812	SAWNGWSLDWSPFGGQEVTRMNAEHTQAGTSVTARNVGNAGVAPGASVGFQTSRSG	871
Qy	940	-----SFGSVNPATPTTA 951	
Db	872	TNAPEGFVAVAGRACPTA 889	
RESULT 3			
Qy	Q9AQH0	PRELIMINARY; PRT; 996 AA.	
AC	Q9AQH0		
DT	01-JUN-2001 (TremBLrel. 17, Created)		
DT	01-JUN-2001 (TremBLrel. 17, Last sequence update)		
DT	01-OCT-2003 (TremBLrel. 25, Last annotation update)		
DE	Glycosyl hydrolase 5 (Fragment)		
OC	Caldicellulosiruptor sp. Tok7B.1.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;		
OC	Caldicellulosiruptor.		
OX	NCBI_TaxID=80339;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
EX	MEDLINE=20171169; PubMed=10706665;		
RA	Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,		
RA	Bergquist P.L.;		
RT	"Multidomain and multifunctional glycosyl hydrolases from the extreme		
RT	thermophile Caldicellulosiruptor isolate Tok7B.1.";		
RL	Curr. Microbiol. 40:333-340(2000).		
DR	EMBL; AF078038; AA06388.1; -.		
DR	HSSP; Q06851; INBC.		
DR	GO; GO:0016787; F:hydrolase activity; IEA.		
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.		
DR	InterPro; IPR001956; CBD 3.		
DR	InterPro; IPR008965; Cellul bind.		
DR	InterPro; IPR002860; Glyco_Hydro_BNR.		
DR	InterPro; IPR011040; Sialidase.		
DR	Pfam; PF02012; BNR; 9.		
DR	Pfam; PF00942; CBM 3; 1.		
DR	ProDom; PD001947; CBD 3; 1.		
KW	Hydrolase.		
FT	NON_TER		
SQ	SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;		
Query Match 47.1%; Score 2420.5; DB 2; Length 996;			
Best Local Similarity 48.5%; Pred. No. 5.2e-94;			
Matches 462; Conservative 151; Mismatches 304; Indels 35; Gaps 15;			
Qy	14	RLVSLAATASFAVAALGLVLPATITASPAHAAT---TQPYTWSNVAI-GGGGFDGIV 69	

Db	2	KELISL-----FSILLILFLIAEFLQIGAKADTGIAVQPVYKMKVIEGGGFIQIV	55
Qy	70	FNEGAPGILYVTRDIDGGMYRWDAANGRWIPLLDWVGNNWNGVGVVSIADRPINTNKV-W	128
Db	56	FNPKEKLVVYRDILOGAYRSTDDGNTWTQLMDWSFDDNLLGVESDIATDPVDPNKKVIL	115
Qy	129	AAVGMYTNSWDPNDAILRSSDQATWQITPLPFLKLGNNMPGRGMGERLAVDPNNNDILY	188
Db	116	ACQGYTNSWTDNMGAILLASTDEGDTFEITPLPFLXGGNMPXRNLERLAIDPNNRILY	175
Qy	189	FGAPSGKGLWRSTDGATWSQMTNPPDVGTIYANPTDITGYSQSDIOGVVWVAADKSSSL	248
Db	176	LCTREGXGLWKSIEDYGVSWKVTSPFNPGTYIEDPNCNDYLNHITGVVWVDFDTSRGP	235
Qy	249	QOASXTIFGVADPNPNFWSRDGGATWQAVPGAPTGFPHKGVDFDPNHVLVIATSNTG	308
Db	236	GEGXIIYVGADKTTSIYTKDGGQTWQALPQPTGLLPQRAKLSS-DGMLIITYSNQ	294
Qy	309	GYDSSGVDVWKFSTGVTWTRISVPVSTDTANDYFGYSGLTIDRQHPNTIMVATQISW	368
Db	295	GPYNGDYGEVWRYNTKTGEWKNISPMQAQDT---YFGYGLAVDAQNPVMAALSSW	351
Qy	369	PTIIIFRSDGGATWTRIDWTSYPNRSRLRYLDISAEPWLTFG-VQPNPPVSPKLGWM	427
Db	352	PDYIWRSIDGGETWKCIWEWNGYPNRTLHYNMDSIAAPWLNFGXTBPTPPEVSPKLGWM	411
Qy	428	DEMAIDPNRDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLBEETAUNDLISPPSG	487
Db	412	VGTLEIDPFNSDXMLYGTGAXLYGCDLTDNDWKGQNTIKVKAIGIBETSQALISPPVG	471
Qy	488	APLISALGDLGFTHADVTAVPSTIFTSPVFTTGSTVDYAEALNPSIIVRAGSFPDPSQP	547
Db	472	PHLFSALGDIAGFRHEDLEKAPNWTYVQPNWGTITDIDFAELNPNFVRVGNVDKQNP	531
Qy	548	DRHVAFTDGGKNWFQ-SEPGGVTTGGTVAASADGSRFVWAPDGPQPVVYVGFQNSW	606
Db	532	TNRIGFSYDGGKSWFQNTPEQGTSEGGTVAAADGSAVVWAP--KGAKVCYSTDNGNKW	589
Qy	607	AASQGVANAIQRDRVNPNTFYALSNGTFRSTDGGVTF-OPVAAAGLPSSGAVGVMFHA	665
Db	590	VECANVPSBAIVYSDRVNPNKFYAFKNGKFIYISADKGTIESPAAGLPISG----NPKT	645
Qy	666	VPKEGDLAASSGLYHSTNGSSNSAITGVSSAVNVGFGKSAPGSSYPAVFVVGTTGG	725
Db	646	VPGIEDIWLVGNNGMWSTDGGYSFKISGVEDAASIGFGKPAEGTYTFAIYYAKING	705
Qy	726	VTGAVRSDDCGTTWVLINDDHOQYGNWGOAITGDHANLRVYIGTNGRGIYDGGAPS	785
Db	706	VRGIFRSDDCDCTWIRINDDKHQFGCANADITGDPVYGVFVATNGLIKWGEIA---Y	762
Qy	786	GSFSP	845
Db	763	SNISPSATPTSTPT	822
Qy	846	PSSSP	904
Db	823	MTPTPTP-----TPTPTPTGTGSLKVLKYNKNETSASAGSIRPFWKLVNCGSSVD	876
Qy	905	LSVTVTYRWFTRDGGSTLVYNCDDWAAIGCGNIRASFSGSVNPATPTADTYIQ	956
Db	877	LSRVKIRYTWTVDDGKPOQSAV-CDWAIQIGASNTFNFVKLSSGSGADYYLE	927
RESULT 4			
Qy	Q97KK0	PRELIMINARY; PRT; 839 AA.	
ID	Q97KK0		
AC	Q97KK0		
DT	01-OCT-2001 (TremBLrel. 18, Created)		
DT	01-OCT-2001 (TremBLrel. 18, Last sequence update)		
DT	01-MAR-2004 (TremBLrel. 26, Last annotation update)		
DE	Probably secreted sialidase; several Asp-boxes and dockerin		
DE	domain.		


```
GN OrderedLocusNames=CAC0919;
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA DOI=10.1128/JB.183.16.4823-4838.2001;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Giu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L.A., Soucaille P.,
RA Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838 (2001).
RL EMBL; AE007608; AAK78895.1; -.
DR PIR; D97013; D97013.
DR HSSP; P38686; 1DAV.
DR GO; GO:000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002860; Glyco hydro BNR.
DR InterPro; IPR011041; Quino gluc DH.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00404; Dockerin_1; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW Complete proteome.
SQ SEQUENCE 839 AA; 90824 MW; 6C99A041CA9CF984 CRC64;

Query Match 39.7%; Score 2037; DB 2; Length 839;
Best Local Similarity 49.5%; Pred. No. 5.7e-78;
Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;

QY 13 SRRVLSLAA-----TASFAVAALGVLPPIAITASPAHAAT--TPYTWNSVAIGGGFVD 66
DB 2 NKRIVSMVAGLSIITTFG-----VTHISAAANKAAASVQSQYKWNDAKIGAGGVP 54
QY 67 GIVNEGAPGILYVRTDIGMYRMDAANGRIPLDDWGNWNGYGVVSYAADPINTNK 126
DB 55 AVIFNKTEKOLIARTDMGGAYRWDAKANKWIPITD--GFSWTMLGCESIATPIDNR 112
QY 127 VAAAVGMYTNSWDNDGAILRSSDQATWQITPPLPKLGGNMPGRGMRERLAVDNNNI 186
DB 113 VYIAAGLYTNDQDENAYILSQDKGNTWKRYQLPFKVGNGNMPGRNGERLQIDPNDKI 172
QY 187 LYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIOGVVWVAFDKSS 246
DB 173 LYLGAARSGNGLWKSSEDYGGTWSKVDNFPDPTGDYQDPQNE--YTADKVGWVWETFDPSGT 230
QY 247 SILGQAKTIFVGVAD--PNPNFWSRDGATWQAVPGAPTGFPHKGVPDPVNHVLIYATS 305
DB 231 TKGSPQTQMYGAADKTGNNIYVNDGKTSWAVKQPKGYLPHGIL-ASDGLYISVS 289
QY 306 NTGGPYDGGSDGVKVPSTGWTTRISVPSTDTANDYFGYSGLTIDRQHPNTIMVATQI 365
DB 290 NTGPGYDGGSDGVKVPSTGWTTRITPPAVGDTKS---GFGISVDAQNPNNVVVATLN 346
QY 366 SWPPTIIFRSTDGATWTRIDWTSYPNRSIRYVLDISABPWLTFG-VQPNPPVPSKL 424
DB 347 RWPDPDEIYRSTDAGTKWPIWMNGYPNRTLYGYNLDYSAQWLDWGKGTGVTTPDPLVKL 406
QY 425 GWMDKMAIDPNSDRMLYGTGATLYATNDLTWKDSGGQIHIAPMVKGLEETAVNDLISP 484
DB 407 GMMGDLEIDPNSDRMFYGTGATLYGTDLLNWDKGNVDLSVKAANGIEECANDVVVP 466
QY 485 PSGAPLISALGDLGFTHADVTAVPSTIFTSPVFTTGTSDVYAEINPSIIVRAGSFPSS 544
DB 467 TKGALLSAVGDDCGFYHDDITKVPSEKMTTPNFSATTISIDYAESVPEFVVRGVGVDTSK 526
QY 545 QPNDRHRAFSTDDGGKNWFQ--GSEPGGVTTGGTVAASADGSRFVWAPGDPGQPVVAVGFG 603
```

```
DB 527 NQODKDCGISYDGGKWNFSAGSNIISGVYKAGTVAAGADAKTIVWSP--EEGANAAYSTDNG 585
QY 604 NSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFFYSTDGGVTF-QPVAAGLPSSGAVGM 662
DB 586 NKWTPCSGLPQAKVRSRVNPKPFYGLNGKFIYSTDAGATFTOSSQTGLPKGK-GI- 643
QY 663 FHAVPGKEGDLWLA--ASSGLVHSTNGSGSWSAITGVSSAVNVNFGKSAPESSYPVAVFVG 721
DB 644 PRTVIGHEGDIWIAGKDLWHSTDSGATFTKVGSDASDTVLGKSKTDDGYPALYMDA 703
QY 722 TIGGVTGAYRSDCCTTWWLINDDQHYGNWGOAITGDHANLRVRYVINGRIVYGDIG 781
DB 704 TIDGTAGIFRSDDEGATWVRINDDAHQYSPDYCITGDPNKYGRVFGVNGRIGVYGDID 763
QY 782 GAPSGSPSPSVS 793
DB 764 GS-QPTPTPSVT 774

RESULT 5
Q70DK5 PRELIMINARY; PRT; 842 AA.
ID Q70DK5; AC Q70DK5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Beta-1,4-xyloglucan hydrolase.
GN Name=xghA;
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN RP SEQUENCE FROM N.A.
RC STRAIN=F7;
RA Zverlov V.V.;
RL Submitted (OCU-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ585344; CAB51306.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR PROSITE; PS00448; CLOS_CELLULOSOME_RPT; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW Hydrolase.
SQ SEQUENCE 842 AA; 92393 MW; CC686932DBA95FD3 CRC64;

Query Match 39.1%; Score 2007; DB 2; Length 842;
Best Local Similarity 50.8%; Pred. No. 1e-76;
Matches 399; Conservative 107; Mismatches 250; Indels 30; Gaps 16;

QY 14 RRLVLLAATASFAVAALGVLPPIAITASPAHAATTPYTWNSVAT--GGGFGVDGIVFNE 72
DB 3 KKFTSKIKAAVFAAVAAATAIFGPAIS---SQAVTSVPYKNDVNVVGGGGGFGPGIVFNE 59
QY 73 GAPGILYVTRTDIGMYRMDAANGRIPLDDWGNWNGYGVVSYAADPINTNKWAAVG 132
DB 60 TEKDLIYARADIGAYRWDPSTETWIPLLDHFQMDYSYGVESIAITDPVDRVYIVAG 119
QY 133 MYTNSWDPNDAILRSSDQATWQITPPLPKLGGNMPGRGMRERLAVDPNNDNILEFGAP 192
DB 120 MYTNDWLPNWGAILRSTDGRTGRTWETKILFPKMGNGPGRSMGERLAIDPNDNRILYLGR 179
QY 193 SKGLWRSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIOGVVWVAFPKSSSSLGQAS 252
DB 180 CGNGLWRSTDGTYGVTWSKVESFPNPGTYIYDP--NFDYTKDIIIGVWVVFDPKSSSTPGNPT 237
QY 253 KTIIFGVADPNPNFWSRDGATWQAVPGAPTGFPHKGVPDPVNHVLIYATNTGPGYD 312
```


Db 298 SGVWTDISP---TSLSAYFYGGSLVDLPVPGILMVAALNCWPEDELIIFRSTDSGATWS 354
QY 385 RIWDTSPNRLRYLVDLSAEPMLTFGVQNP-PVPSPKLGWMDMAIDPFPNSDRMLY 443
Db 355 PIWENGYPSINYYSYDISNAPMTQTTSDQFV---RVGMVVEALAIIDPFDNSHWLY 411
QY 444 GTGATLYATNDLTWDSGGQIHAPMVKGLBETAVNDLISPSGAPLISALDGLGFTHA 503
Db 412 GTGLTVYGGHDLTNWDSKHNVTVKSLAVGIBEMAVLGLITPPGGPALLSAGVDDGGFTHS 471
QY 504 DVTAVPSTFTSPVFTTGSVDYAEINPSIIVRAGSPDPSSOPNDRHVAFTSDGCKNWFQ 563
Db 472 DDAAPNQAQYHTPYGTNGIDYAGNKPENIVRSGAD--DYPT---LALSNFSGTWYA 526
QY 564 GSEPGVTTGGTVAASADSRFWAPDGPQPVVYAVFGNSWAASQGVPAQAQIRSDRV 623
Db 527 DYAASTSTGTGAVALSADGDTVLLMSSTSGALVSKSQG---TLTAVSSLPSCGAVIASDKS 583
QY 624 NPKTYVALSNGTFYRSTGGTTFQVAAGLPSSGAVGVNHFHAPVKEGDLMLAASGLYH 683
Db 584 DNTVYFGSAGAIYVSKNTATSTFTKVS-LGSSTTVNAI-RAHPSIAGDVMASTDKGLWH 641
QY 684 STINGSSWSAI-TGVSSAVNVGFGKSPGSSYPVAVVVGTTIGVTVGAYRSDCGTTWVLI 742
Db 642 STDYGSTFTQIGSGVTAGSFGKASSTGVYIYGFPTIDGAAGLKFSEADAGTNQVY 701
QY 743 NDDQHQYGNWQA---ITGDHANLRVYIGTNGRGIVYGDIGGAPSG-----SPSPSV 792
Db 702 SDASHFGS-GSANVNGDLQYGRVFRGHERPGHLLRQSGREPAGHGDGDDTTSTKT 760
QY 793 SPASPSLSPPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP 852
Db 761 STTVSTLTKTSSASTSSSTTVKTTTSSSTTSKASSTTTTKTTTSSSTGTTATA 820
QY 853 SS 854
Db 821 SA 822

RESULT 9
Q729M8 PRELIMINARY; PRT; 838 AA.
AC Q729M8; (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cel74a.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QM6a;
RA Foreman P.K., Brown D., Dankmeyer L., Dean R., Diener S.,
RA Dunn-Coleman N.S., Goedegebuur P., Houfek T.D., England G.J.,
RA Kelley A.S., Meerman H.J., Mitchell T., Mitchinson C., Olivares H.A.,
RA Teunissen P.J.M., Yao J.J., Ward M.;
RT "Transcriptional Regulation of Biomass-Degrading Enzymes in the
RT Filamentous Fungus Trichoderma reesei.";
RL J. Biol. Chem. 0:0-0(2003).
RL EMBL: AY281371; AAP57752.1; .
DR GO: 0005576; C:extracellular; IEA.
DR GO: 0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO: 0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR00254; CBD_fungal.
DR InterPro: IPR002860; Glyco hydro_BNR.
DR InterPro: IPR011040; Sialidase.
DR Pfam: PF02012; BNR; 6.
DR Pfam: PF00734; CBM1; 1.
DR ProDom: PD001821; CBD_fungal; 1.

DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
SQ SEQUENCE 838 AA; 87132 MW; FBE3D08F2A74FA84 CRC64;
Query Match 31.7%; Score 1625.5; DB 2; Length 838;
Beat Local Similarity 41.7%; Pred. No. 1.1e-60;
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;
QY 25 SFAVAAALGVLPATIASPAHAATTPQYTSNVAI--GGGGFVDGIVFNEGAPGLIYVRTD 83
Db 4 SRLVALVGA-----IPAHAA-----FSWKVNVKLGCGGFGVPGIIFHPKTKGVAYARTD 53
QY 84 ICGMYRWDAAANGRWIPLLDW-----GWNVNGVGVVSIADPINTNKVMAAVGMYTNSWD 139
Db 54 IGLYRLN-ADDSWTAVTDDGIADNAGHNW---GIDAVALDPQDDQKVAAVAGMYTNSWD 109
QY 140 PNDGAILRSSDGCATWQITPLPFKLGNNMPGRGMBERLAVDPNDNLIYFGASGKGLWR 199
Db 110 PENGAIIRSSDRGATWSFTNLPFKVGGNNMPGRGAGERLAVDPANSNIIFYFGARSGNGLWK 169
QY 200 STDGATWSQMTNFPDVGTVIANPTDITGVQSDIQGVVWVAPDKSSSSILGOASKTIIFGV 259
Db 170 STDGGVTFSKVSFTATGYIIPDSDSNGYNSDKQGLMMVTFDSTSTTCGATSRI FVGT 229
QY 260 ADP-NNPVFWSRDGGATQAVPGAPTGFIPHKGVDFVNVHLYIATSNTPGYPDSSGDV 318
Db 230 ANNITASVYVSTNAGSTWSAVFGPKYFPHKAKLQAPAEKALYLYTSDGTGPTGLGSV 289
QY 319 WKFSVTSCTWTRISVPSTDTANDYFGYSLGTLTIDRQHPNTIMVATQISWPDITII FRSTD 378
Db 290 WRDYIAGGTWKDITPVSGSDL---YFGFGGLGLDLQKPGTLVVASLNSWMPDAQLFRSTD 346
QY 379 GGATWTRINDWTSYPNRSLRYVLDLSAEPMLTFG-VQNPFPVS---PKLGWMDMAAI 433
Db 347 SGTWSPPIAWASYPTEYYSISTPKAPWKNNFIDVTSESSESDGLIKRLGMIESLEI 406
QY 434 DPNSDRMLYGTGATLYATNDLTWDSGGQIHAPMVKGLBETAVNDLISPPSGAPLISA 493
Db 407 DPTDSNHWLYGTGWTIFGGHDLTNWTRHNVSQSLADGIEEFVQDLASAPGSELLAA 466
QY 494 LGDLGCFTHA---DVTAVPSTFTSPVFTTGSVDYAEINPSIIVRAGSPDPSSOPNDRH 550
Db 467 VGDDNGFTFASRNDLGTSPQVWATPTWATSTSDYAGNSVKSVMRVGNTAGTQQ----- 521
QY 551 VAFSTGGKNWFOGSEPPGGVTTGTVAAADGSRFWAPDGPQPVVYAVFGNSWAASQ 610
Db 522 VAISSDGGATWSIDYAADTSMNGTVAYSADGDTILWSTASSG---VQRSQFGSFAVS 578
QY 611 GVPANAQIBSDRVNPKTFVALSNGTFFVRSYTDGVTQFVAAGLPSSGAVGVW---FHAVPG 668
Db 579 SLPAGAVIASDKKTNSVYFVAGSGSTFVSKDTGSSP---TRG-PKLGASGATIRIDIAHPT 634
QY 669 KEGDLMLAASSGLYHSTNGSSWSAI-TGVSSAVNVGFGKSPGSSYPVAVVVGTTIGV 727
Db 635 TAGTLVSTDVGI FPRSTDGTTFGQVSTALTNTYQIALGVGS-GSNW-NLVAFTG--GPS 690
QY 728 GA--YRSDDCGTTWVLIINDHQYGNWQQAIGDHANLRVYIGTNGRGIY--GDIGGA 783
Db 691 GARLYASGDSGASWTDIQSQSGFSGIDSTKVAGSGTAGQVYGTNGRGVYAOQTGVGG 750
QY 784 PSQSPSPSVFSPASPSLSPPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 843
Db 751 TGGT-----SSSTKQSSSTSTSSSTTLRSSVVSTTRASTVTSRTSSAAGPTGS 801

RESULT 10
Q7S1W6 PRELIMINARY; PRT; 861 AA.
AC Q7S1W6;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.


```

GN Name=NCU05955.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels P., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamyselis M., Mauceli E., Bielke C., Rudd S., Frisman D.,
RA Krystofova S., Rasmussen C., Metzenberg R.B., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmari S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nisbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
DR EMBL; AAX01000439; EAA29333.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011041; Quino Gluc DH.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00734; CBM_1; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR KX Hypothetical protein.
SQ SEQUENCE 861 AA; 89114 MW; 95BDBE846631C67D CRC64;

```

Query Match 30.9%; Score 1587.5; DB 2; Length 861;
 Best Local Similarity 38.0%; Pred. No. 4.5e-59;
 Matches 356; Conservative 169; Mismatches 306; Indels 107; Gaps 27;

```

QY 11 MRSRLVSLAATASFAVAALGVLPITAITASPAHAATQPVTSNVAI-GGGGFGVGGIV 69
DB 1 MKSISVQLLLAAA-----AAPVRAAA-----SKNWNIGGGGFGVPGFV 40
QY 70 FNEGAGILYVRTDIGMYRWAANGRIWPIPLLDWV-----GWNNGYGVWSIAADPINTN 125
DB 41 FHPTEKGAVARTDIGLYRLN-ADDSMTAITSITDDKNNW---GIDAVALDPQDPD 96
QY 126 KWAAVGMYTNSDPNDGAILRSSDQATWQITLPPFKLGGNMPGRCGERLAVDPNNDN 185
DB 97 KYVAATGMYTNSDPNNGTLIRSSDRGATWSSTLPPFKVGNMPGRCGERLAVDPKNSK 156
QY 186 ILYGAPSGKGLWRSTDSTGATWSQNTNFPDVGTYIANPTDITGYSQIDQGVVWVAFDKSS 245
DB 157 ILYFGARSGHLYKSTDSTGATWSQNTNFPDVGTYIANPTDITGYSQIDQGVVWVAFDKSS 216
QY 246 SSLGQASKTIFVGVADPNPNPVFW-SRDGGATWQAVPGAPTGFIPHKVDFPNVHLYIAT 304
DB 217 ATTNGATSRIFVGATNTSSVWVNSDAGATWSAVAGPGYFFHKCKLQTERALYLYTY 276
QY 305 SNTGPGYDGGSGDVWKFVSVTGWTTRISPVSTDTANDYFGVSLGITDRQHPNTIMVATQ 364
DB 277 SDGTGPGYDGTAGAVRVYDITNGTKDIT--PGSDLA---YGFGLGVDMKNPFGTIVASL 331
QY 365 ISWHPDITIFRSTDGGATWTRIDWTSYPNLSRLVLDISAEPLMTFGVQNPFPVPSKL 424

```

```

Db 332 NLWMPDAQIYRSTDSTGATWSPIWMDGYPNMNQYGLTTPNAPWTETGFLSQ--DTKHL 388
QY 425 GWMDEAMAIIDPFNSDRMLYGTGATLYATNDLTWKDSGGQIHIAIPVMVKGLEETAVNDLISP 484
Db 389 GWMIESLEINFLDSHMLYATGLTVYGGHDLTKWDTVHNVTIQSLAVGLEEWAIVGLASA 448
QY 485 PSGAPLISALGDLGGFT---HADVTAVPSTIPTSPTFTTGTSTVDVYAEALNPSIIVRAGSFD 541
Db 449 PGSELLAAVGDGCGFTFKSSDLGTSKPTPMWTPQWASSADVDYAGNKPANVVRIGSGS 508
QY 542 PSSQPNDRHVAFTDGGKWFQGGSPGGVTTCGTTVAASADGSRFWAPGDPGQPVVAVG 601
Db 509 GAQQ-----VAVSSDGGASWHAHNGTDTTKSGTVAYSADATIWMSSGTSG--VVRSQN 561
QY 602 FCNSWAASQGVPAANAIRSDRVNPKTFYALS-NGTFYRSTDCGVTPOPPVAAGLPSSGAV- 659
Db 562 QG-IPTAVASIPSGAVIASDKRNTVTFAGSTTGTFYRSTDTGATFTAVSDALGSAKAVR 620
QY 660 GVMFHAVPGKEGDLMLAAASSGLYHSTNGGSSWSAITGVSSAVNVGFGKSPGSSYPVAFV 719
Db 621 DIVAH--PTVAGELYVSTDAGIFRSTDFGVSTFKLSGLTNTQSVLSGVGSTSGW-NLYA 677
QY 720 VGTIGGVTVGVRSDCGTTWVLINDDQHQYGNWQA---ITGDHANLRRVYICTNGRGIV 776
Db 678 FGTGANGKLYASADAGATWADVQGAQ-GFGSIAVASSKVAGSANVPQGVYVGTNGRGVL 736
QY 777 YCDIGGAPSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 836
Db 737 YAQV--SVSGT-----APGAT-----SSSAPVKTSSTAAAPSSAAVAKTSSTS 780
QY 837 SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 896
Db 781 AAAPVSITTSKAAATTTLVSTVKVPTATLTS-----SVK 817
QY 897 NTGSSVLDSTVTVRYWFTRDG---GSSTLV--YNCD 928
Db 818 STSTVGAEIA---QHWYQCGGVGTGPTACVSPYKCE 851

```

RESULT 11
 Q82K30
 ID Q82K30 PRELIMINARY; PRT; 739 AA.
 AC Q82K30;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative glycosyl hydrolase.
 GN OrderedLocusNames=SAV2574;
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RC STRAIN=MA-4680;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 avermitilis: deducing the ability of producing secondary
 metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

Db	508	QFNLDSIDAAGNFPNVVRAGSSGHEYDSACARGAYATDGDWTFPTCPFGMNASHY	567
Qy	573	-GGTVAASADGSRFVWAP--GDFGQPVVYAVGFCNSWAASQGYPA-----NAQIRSDRV	623
Db	568	QGSTIAVDASGSQIWMSTKLDEQASGWPYSHDYKGTWS----VPAGDLKQATANVLSDKV	623
Qy	624	NPKTFYALSNGTFRSTDDGGVTQFVAAAGLPSSGAVGVMFHAVPGKEGDLWL-AASSGLY	682
Db	624	QDGTFTYATDGGKFFVSTDGKSYAAKAGGLVT--GTSLMPAVNPWVAGDVWVPVPEGGLF	681
Qy	683	HSTNGSSWSAI-TGVSSAYNVGFGKS---APGSSYPVAVFVGT--IGGVTGAYRSDDC	735
Db	682	HSTDGASFTRVGTANATLVSGAPKSKSDGKKASAPSAVFIWGTDKPGSDIGLYRSDDN	741
Qy	736	GTTWVLINDQHQYGNWGOAITGDHANLRVYIGTNGRGIVYDI	780
Db	742	GSTWTRVNDQEHNYSG-PTWIEADPKVYGRVYLGVTNGRGIVYADL	785

Search completed: October 5, 2005, 07:56:03
Job time : 124.987 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:49:28 ; Search time 36.7684 Seconds
(without alignments)
1942.949 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLTWRRRLVSL.....RASPGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgm2_6/ptodata/1/iaa/5A COMB.pcp:*
2: /cgm2_6/ptodata/1/iaa/5B COMB.pcp:*
3: /cgm2_6/ptodata/1/iaa/6A COMB.pcp:*
4: /cgm2_6/ptodata/1/iaa/6B COMB.pcp:*
5: /cgm2_6/ptodata/1/iaa/PCITUS COMB.pcp:*
6: /cgm2_6/ptodata/1/iaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	334	6.5	1751	3	US-09-136-574A-44
2	301.5	5.9	1749	4	US-09-640-419C-28
3	288	5.6	96	4	US-09-119-507B-112
4	288	5.6	96	4	US-09-547-693-236
5	276	5.4	1426	3	US-09-136-574A-43
6	260	5.1	2736	4	US-09-252-991A-30227
7	234	4.6	2137	3	US-09-134-001C-4463
8	225.5	4.4	551	2	US-09-033-537A-1
9	216.5	4.2	8991	4	US-08-714-741-32
10	215.5	4.2	76	4	US-09-547-693-233
11	213	4.1	3892	4	US-09-328-352-5503
12	210.5	4.1	521	1	US-08-276-213-3
13	208.5	4.1	206	4	US-08-529-055-54
14	208.5	4.1	493	3	US-09-198-956-10
15	208.5	4.1	493	3	US-09-198-955A-12
16	208.5	4.1	493	3	US-09-694-531-12
17	208.5	4.1	493	4	US-09-670-141-10
18	208.5	4.1	493	4	US-10-072-152-12
19	205	4.0	918	4	US-09-200-650E-1
20	204.5	4.0	1970	4	US-09-538-092-1005
21	202.5	3.9	476	4	US-09-339-159B-4
22	200.5	3.9	700	2	US-07-862-588B-2
23	200	3.9	183	4	US-08-529-055-50
24	199.5	3.9	490	3	US-09-109-841-2
25	199.5	3.9	616	3	US-09-136-574A-47
26	197	3.8	1060	3	US-08-911-393-2
27	197	3.8	1060	4	US-09-955-909-2

28	192.5	3.7	271	4	US-09-248-796A-21050	Sequence 21050, A
29	191.5	3.7	412	1	US-08-313-288B-18	Sequence 18, Appl
30	191	3.7	105	4	US-09-547-693-230	Sequence 230, App
31	191	3.7	2169	4	US-09-949-016-6930	Sequence 6930, Ap
32	190	3.7	1092	4	US-09-147-405B-15	Sequence 15, Appl
33	189.5	3.7	1187	4	US-09-949-016-6513	Sequence 6513, Ap
34	189	3.7	933	3	US-08-293-728-2	Sequence 2, Appl1
35	189	3.7	933	3	US-09-421-868-2	Sequence 2, Appl1
36	188.5	3.7	193	4	US-08-529-055-49	Sequence 49, Appl
37	187	3.6	936	4	US-08-956-171E-4249	Sequence 5249, Ap
38	187	3.6	936	4	US-08-781-986A-5249	Sequence 235, App
39	186.5	3.6	104	4	US-09-547-693-235	Sequence 1, Appl1
40	186	3.6	423	2	US-08-760-797A-1	Sequence 1, Appl1
41	185.5	3.6	424	3	US-08-932-929B-1	Sequence 42801, A
42	185.5	3.6	793	4	US-09-270-767-42801	Sequence 4, Appli
43	185.5	3.6	1719	2	US-08-459-568-4	Sequence 4, Appli
44	185.5	3.6	1719	2	US-08-399-411-4	Sequence 4, Appli
45	185.5	3.6	1719	3	US-08-516-859A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1751 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match	6.58; Score 334; DB 3; Length 1751;
Best Local Similarity	21.24; Pred. No. 8e-13;
Matches	219; Conservative 125; Mismatches 314; Indels 376; Gaps 56;
Qy	18 SLLAATASPAVAALGLVPLAITASPAHAA-----TTQP---YTWNSVAI 59
Db	186 SVVAET-----AASLAVASVVIKERNISQKAASYLOHAKDLPEFADTTRSDAGYTAAGTFY 240
Qy	60 GGGGFVDGIVFNBCGAPGILYVRTD-----IGCMYRW-----DAANGRWIP 99
Db	241 TSGGFIDDLGW--AAVWLYIATNDSSYLTKABELMSEYANGTWTWTCQWDDVRYGTLIM 297
Qy	100 LLDWVCWNNKGYNGVVSIAADPNTNKVWAAGVM-YTNSWDPNPDGAILRSSDQG-----152
Db	298 LAKITKEL--YKGAVERNIDHWTDRIITYTPKGMAYLTGW---GSLRYATAAFLACVY 351
Qy	153 ATWQ-----ITPLPKLGNMGR---GMGERLAVDPNNDNILYFGAPSG 194
Db	352 ADMGCGDSNKKTKYLNPAKSQIDYALGST--GRSFVVGFGYNYQHPHNR-----400
Qy	195 KGLWRSDSGATWSQMTNFPD-----VGTVIAMPTDTTQSDIQGVWV--VAFDKSSS 246
Db	401 -----AHSSWANSMKIPEYHRIILYGALVCGPGSDDSINDJITDYQNEVACDYNAG 452
Qy	247 SLGOAKTIIFVGVADPNPNFWSRDKGATQWAVPGAPTGFPHKGHGFDPVNNHLYIAT--304
Db	453 IVGALAK-----MYQLYGGEPID--DFKAIETPTNDEIFVESKF 489
Qy	305 SNTGGPYDSSGDVWKSPTSVTSGTWTRISPPVSTDTAN-DYE-----GYSGLTIDRQ 354
Db	490 GNSQGP---NYTEVISYIYNRTGW-----PRRVTDKLSFKFYIDITELIQAGYS-----535
Qy	355 HPNTIMVATQISWMPDRTIIFRSTGGGATWRIWDWTYPNRSRLRYL-DISAEPLWTFGV 413
Db	536 -PDVVKDT-----YYIEGKLSGPYVWD---KRNRIYVVLVDFSGTK-----I 575
Qy	414 QPNPPVPSPKLGW---MDEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPM 469
Db	576 YPGGEVEHKQAOQFKISVPGYPWDPTN-DPSYKGLTSQLEKNKYIAAYDNN-----L 628
Qy	470 VKGLETAVNDLLSP-----PSGAPLISALGDLGGFTHADVTAVPSTIIT---SPVFT 519
Db	629 VWGLEPGAATSTAPTSTPTPTPTPTVTA-----TPTPTPTPTPTGSP--G 674
Qy	520 TGTSDVYAEINPSIIVRAGSFDPSQPNDRHVAFTDGGKNWFGSGSEPGVTTGCTVAAS 579
Db	675 TSGGVKVLKYNNETSASTGSIKP-----WFK-----IVNGG---SSS 708
Qy	580 ADGSRF---VM--APGDPGPVVYAVGVFGNSWAASQGVPAQAQIRSDRPNKPTFVLSNG 634
Db	709 VDLSRVKIRYWTYVDGDKPQSAVC-----DM-----AQIGASNVT-FNFVKLSSG 752
Qy	635 TFRSTDGGVTPOVAAAGLPSGGAGVGMFHAVPGKB-GDLMLAASSGLYHSTNGSGSSWA 693
Db	753 V-----SGADYY-LEVGF-SSGAGQLQ---PGKDTGDIQVRFNKNDWSNYNQADDWSW 800
Qy	694 ITGVSSAVNVGPKSAPGSSYPVAVFVGTGTGVTGAYRSDCGTWTWILNDHQHQYGNWG 753
Db	801 LQSMTN-----YGENAKVTLY-----VDG-----VLV-----WG 824
Qy	754 QAITGDHANLRVYIGTNGRGIVYDGIIGAPSGSPSPSPSPSPSPSPSPSPSPSPSPSPSP 813
Db	825 Q-----EPGGA-----TPAPTSTATPT 841
Qy	814 PPSRSPSSP 873
Db	842 PTPATPTPTPTPTPTVTSATPTPTAPTASPVGSGYWTSES-----YCAL 887
Qy	874 KVQYKNDSAPGDNIQIKPGLQVNTGSSVDLSVTYRYWTRDGGSTLVYNCDDWAIG 933
Db	888 KVVYANGLSPTNVLNPKIKIENVGTAVDLRSVKRYWYTTIDGCAATQS-----938

```

QY      934  CGNIRASFGSVNPA 947
      :
      :|||
Db      939  -----SVASSINPA 947

RESULT 2
US-09-640-419C-28
; Sequence 28, Application US/09640419C
; Patent No. 6630615
; GENERAL INFORMATION:
; APPLICANT: Bidney, Dennis L
; APPLICANT: Crasta, Oswald R
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; TITLE OF INVENTION: DEFENSE-RELATED SIGNALING GENES AND METHODS OF USE
; FILE REFERENCE: 35718/199009 (5718-92)
; CURRENT APPLICATION NUMBER: US/09/640,419C
; CURRENT FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,656
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/206,405
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1749
; TYPE: PRT
; ORGANISM: Synchocystis PCC6803
US-09-640-419C-28

```

Query Match	5.9%;	Score	301.5;	DB	4;	Length	1749;
Best Local Similarity	19.0%;	Pred.	No. 9.7e-11;				
Matches	225;	Conservative	117;	Mismatches	402;	Indels	439;
Gaps	46;						
Qy	17	VSLAATASPAVAALGVLP	IAITASPAHAATTPQY	TWSNVAICGG	-----	62	
		:::::	:::::	:::::	:::::		
Db	23	LAULACSSFSG	---NVLAQNITPA	PDGTGTTVDAQGNQFNIGGSLG	DGQNLPHSLQ	78	
		:::::	:::::	:::::	:::::		
Qy	63			---GFVDGIVFNEG	AGPILVYRTDIGMYEWDA	92	
		:::::	:::::	:::::	:::::		
Db	79	QFLDQGIANPLSNPD	IRNILTRIVGGDASI	INGLIQVSGGNANLFMNPAGMIFG	PNIA	130	
		:::::	:::::	:::::	:::::		
Qy	93	A	-----NGRPIPLD	WGMNNGVGVVS	-----	116	
		:::::	:::::	:::::	:::::		
Db	139	SINVPDGVVTTGSA	IGFGNDQFQV	---FSDNDYNALIGNPSQ	FAFDLANPLGLI	193	
		:::::	:::::	:::::	:::::		
Qy	117	-----IADP	INTNKVAAVGM	YTNSDPNDGAILRSS	-----	149	
		:::::	:::::	:::::	:::::		
Db	194	GDLSVTEGKNLT	FLAGNIVNTGSLA	APGNNITVAAPVQGNRIRISQAGSLLSLE	VEVSPQ	253	
		:::::	:::::	:::::	:::::		
Qy	150	-DQATWQITPLP	FKLGNPNRGRGM	ERLAVDPNND	-----NILYFGAPSGKGLMR	202	
		:::::	:::::	:::::	:::::		
Db	254	MNQGGSFVL	DLPTLLTQ	GASNLDLG	-LAVQNGSVTNGTGNALVSP	FGSVTISG	311
		:::::	:::::	:::::	:::::		
Qy	203	SGATWSQNTFP	PDVGTVIANTPT	TGYQS	---DIQ	-----GVVWAFD	-KSSSSLGQASK
		:::::	:::::	:::::	:::::		
Db	312	AS	---GKSTNISS	-GGVATAGDQIA	VQGAIVDVSGNGGGGT	VRIGDGFQCG	LTLPNASQ
		:::::	:::::	:::::	:::::		
Qy	254	TIFGVGADPNPV	---FWSRDGGA	---TWQAVEGAPT	GFIPHKGVDFPVNHVLIAT	SNT	307
		:::::	:::::	:::::	:::::		
Db	368	TLI	-----DSNSVVKAD	ALLTNGCGT	VIWADDS	TRFSGNI	-----SAQ
		:::::	:::::	:::::	:::::		
Qy	308	GGPYDGS	SGDVWKP	SVTS	SGTWTTRIS	VPSTDTTANDYFGY	SGLIDROHPNTIMVATQISW
		:::::	:::::	:::::	:::::		
Db	407	GGTMGGNGG	---FVETSG	-----AKSLM	VDDTARVNTFATM	GBELGT	444
		:::::	:::::	:::::	:::::		
Qy	368	W	---PDTI	IFRSTDGGAT	WTWIRD	---WTSYPN	BSRLRYVLDISABE
		:::::	:::::	:::::	:::::		
Db	445	WLLDPLEI	IIVGTTDD	LLADPKL	ISVLTIT	ISLONGNVILQ	ADQSI
		:::::	:::::	:::::	:::::		
Qy	421	SPKLGW	MEAWAIDP	PNFSDRMLY	VTGTATLYA	-----TNDLT	KWDSGGQIHI
		:::::	:::::	:::::	:::::		

Db 503 -GNLTDPSTTTIDALFS-----LGTGSIIFANTGPINTGNTLVTSPFTNLDNDKNIQLNA 557

Qy 467 -----APWVGLEETAVN---DLISPPS-----GAPLISALGDLG----- 498

Db 558 NNTTAPGDIYFRKSVNGGFDLGNANFVYFDDGAGITTPKSGFVGTATBIYVGNDIVT 617

Qy 499 -----GFTHADVT-AVPTIPTSPTFTTG-----TSVDYAEILNP 531

Db 618 QGNQIFGVFLGQPVNLTSSAGSVIFTNNTLLNGSLQVTAQNIIVSQPSSSLSAVEIAS 677

Qy 532 SIIVRAGSFDPSSQPNDRHVAF---STDGGRNWFQGSBPGGVTTG-----GTVA 577

Db 678 DVLLNAG-----QNVSEGNINTRGGNVDIQAL--GNISTGSIIVTSFPGGNAGNVI 725

Qy 578 ASADG---SRFVWAPGDPQPVVAVGPGNSWA-----ASQGV 612

Db 726 LNAGGTLTGVYIETSGTNGGVDVTTSSGSNTSTAVIDTRGFGDGLDLSLGGAVSIESKGD 785

Qy 613 PANAQIRSDRVNPKTFYALSNGTFRSTGCVTFQPVAAAGLPSSGAVGVMFHAVPGKE-- 670

Db 786 ITTAFIDTGAYSIESFNGTGNVFLTAGDSITTYIIFTAGKNGG--DIFTOAGESIIL 843

Qy 671 -----GDLWLAA---SSGLYHSTNGGS-----SWSAITGV 697

Db 844 DYLNTYGSQTSQDVVVEAPLDISGSYIYTGGEPEGNVFLQAGDITTSYIDTSAANGG 903

Qy 698 SSANVVG-----FGKAPGSSYPVAVVVGT 722

Db 904 DIFIQSGDTEVGYLFTKGYEGRGVDVYVETGRYFRAIDGFLGEGPFSVVTAGLTVG- 962

Qy 723 IGVVTGAYRSDCGTTWVLINDDQHOYGNWGOAITGDHANLRVYIGTNGR--GIVYGD 780

Db 963 -----GSVYI-----QFGSRPFPIGNPI-----TNGTIGAISSGDD 994

Qy 781 CGAPSGSP-----SPSVSPASPSLSRSPSPSSSPSPSPSPSPSSSPSSSPS 825

Db 995 NTVPIGTPIIDFTFLDNITITTEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 1054

Qy 826 PSP 868

Db 1055 PE 1097

RESULT 3

US-09-119-507B-112

; Sequence 112, Application US/09119507B

; Patent No. 6548642

; GENERAL INFORMATION:

; APPLICANT: Kieliszewski, Marcia J.

; TITLE OF INVENTION: No. 6548642e1 Synthetic Genes for Plant Gums

; FILE REFERENCE: OHU-03417

; CURRENT APPLICATION NUMBER: US/09/119,507B

; CURRENT FILING DATE: 1998-07-20

; NUMBER OF SEQ ID NOS: 118

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 112

; LENGTH: 96

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic

US-09-119-507B-112

Query Match 5.6%; Score 288; DB 4; Length 96;

Best Local Similarity 77.0%; Pred. No. 1.6e-11;

Matches 57; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

Qy 791 SVSPSPASPSLSRSP 850

Db 20 SLSLAQTTRASP 79

Qy 851 SPSSSPSPSPSPSPPT 864

Db 80 SPSPSPSPSPSPSP 93

RESULT 4

US-09-547-693-236

; Sequence 236, Application US/09547693

; Patent No. 6639050

; GENERAL INFORMATION:

; APPLICANT: Kieliszewski, Marcia

; TITLE OF INVENTION: Synthetic Genes for Plant Gums and Other Hydroxyproline-Rich

; FILE REFERENCE: OHU-04089

; CURRENT APPLICATION NUMBER: US/09/547,693

; CURRENT FILING DATE: 2000-04-12

; NUMBER OF SEQ ID NOS: 236

; SOFTWARE: Patent In version 3.0

; SEQ ID NO 236

; LENGTH: 96

; TYPE: PRT

; ORGANISM: Artificial/Unknown

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: Synthetic

US-09-547-693-236

Query Match 5.6%; Score 288; DB 4; Length 96;

Best Local Similarity 77.0%; Pred. No. 1.6e-11;

Matches 57; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

Qy 791 SVSPSPASPSLSRSP 850

Db 20 SLSLAQTTRASP 79

Qy 851 SPSSSPSPSPSPPT 864

Db 80 SPSPSPSPSPSPSP 93

RESULT 5

US-09-136-574A-43

; Sequence 43, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:52:23 ; Search time 121.54 Seconds
(without alignments)
3268.602 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENIRLTMSRRRLVSL...RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues

Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/FCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	957	11	US-09-917-376-1
2	5134	100.0	957	14	US-10-155-400-1
3	4036	78.6	740	11	US-09-917-376-3
4	4036	78.6	740	11	US-09-917-376-6
5	4036	78.6	740	14	US-10-155-400-3
6	4036	78.6	740	14	US-10-155-400-6
7	2478	48.3	882	14	US-10-156-761-9395
8	1680	32.7	726	11	US-09-917-376-7
9	1680	32.7	726	14	US-10-155-400-7
10	1625.5	31.7	838	15	US-10-420-191-2
11	1604	31.2	818	14	US-10-026-994-2

12	1442	28.1	739	14	US-10-156-761-10111	Sequence 10111, A
13	1125	21.9	812	15	US-10-395-241-12	Sequence 12, Appl
14	1117	21.8	789	15	US-10-395-241-14	Sequence 14, Appl
15	1114	21.7	826	15	US-10-395-241-18	Sequence 18, Appl
16	848	16.5	555	10	US-09-927-827-47	Sequence 47, Appl
17	823	16.0	1228	10	US-09-917-384-1	Sequence 1, Appl
18	823	16.0	1228	10	US-09-917-383-1	Sequence 1, Appl
19	702	13.7	762	10	US-09-917-378-1	Sequence 1, Appl
20	511.5	10.0	1043	10	US-09-917-384-6	Sequence 6, Appl
21	511.5	10.0	1043	10	US-09-917-383-6	Sequence 6, Appl
22	462	9.0	88	11	US-09-917-376-5	Sequence 5, Appl
23	462	9.0	88	14	US-10-155-400-5	Sequence 5, Appl
24	462	9.0	89	11	US-09-917-376-4	Sequence 4, Appl
25	462	9.0	89	14	US-10-155-400-4	Sequence 4, Appl
26	462	9.0	154	10	US-09-917-378-4	Sequence 4, Appl
27	459	8.9	150	10	US-09-917-384-5	Sequence 5, Appl
28	459	8.9	150	10	US-09-917-383-5	Sequence 5, Appl
29	389	7.6	284	17	US-10-418-032-270	Sequence 270, App
30	346.5	6.7	901	17	US-10-939-262-12	Sequence 12, Appl
31	336.5	6.6	741	14	US-10-156-761-8100	Sequence 8100, Ap
32	326.5	6.4	628	17	US-10-939-262-14	Sequence 14, Appl
33	288	5.6	96	15	US-10-437-708-236	Sequence 236, App
34	288	5.6	96	15	US-10-395-402-112	Sequence 112, App
35	288	5.6	96	16	US-10-257-199-236	Sequence 236, App
36	288	5.6	96	17	US-10-418-032-236	Sequence 236, App
37	274	5.3	406	18	US-10-450-763-57609	Sequence 57609, A
38	268	5.2	284	17	US-10-418-032-271	Sequence 271, App
39	262	5.1	2468	14	US-10-248-330-4	Sequence 4, Appl
40	262	5.1	2468	15	US-10-282-122A-66335	Sequence 66335, A
41	261.5	5.1	599	10	US-09-955-555A-29	Sequence 29, Appl
42	253	4.9	2435	15	US-10-282-122A-47453	Sequence 47453, A
43	252.5	4.9	2117	15	US-10-120-801-63	Sequence 63, Appl
44	246	4.8	1066	16	US-10-647-195-43	Sequence 43, Appl
45	243.5	4.7	1049	15	US-10-282-122A-49900	Sequence 49900, A

ALIGNMENTS

RESULT 1
US-09-917-376-1
; Sequence 1, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
US-09-917-376-1

Query Match 100.0%; Score 5134; DB 11; Length 957;
Best Local Similarity 100.0%; Pred. No. 3.9e-280;
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDRSENIRLTMSRRRLVSLAATAAFVAAALGVLPITAITASPAHAATTPQYTWNSVAIG 60
DB 1 MDRSENIRLTMSRRRLVSLAATAAFVAAALGVLPITAITASPAHAATTPQYTWNSVAIG 60


```
QY 61 GGGFVDGIVFNEGAGILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 120
Db 61 GGGFVDGIVFNEGAGILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 120
QY 121 PINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLGNNMGRGERLAVD 180
Db 121 PINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLGNNMGRGERLAVD 180
QY 181 PNNDNILYFGAPSGKGLWRSDGATWSQMTNFPDVGTYIANPTDTTGYQSDIOGVVWA 240
Db 181 PNNDNILYFGAPSGKGLWRSDGATWSQMTNFPDVGTYIANPTDTTGYQSDIOGVVWA 240
QY 241 FDKSSSSILGQASKTIFVGVADPNPNVFWSRDGGATWQAVPGAPTGFIPHKGVFDPVNVHL 300
Db 241 FDKSSSSILGQASKTIFVGVADPNPNVFWSRDGGATWQAVPGAPTGFIPHKGVFDPVNVHL 300
QY 301 YIATSNTPGPGYDSSGDVWKFVSVTGWTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360
Db 301 YIATSNTPGPGYDSSGDVWKFVSVTGWTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360
QY 361 VATQISWMPDITIIFRSTDGGATWTRIDWNTSYPNRSRYVLDISAEPLWTFGVQPNPVP 420
Db 361 VATQISWMPDITIIFRSTDGGATWTRIDWNTSYPNRSRYVLDISAEPLWTFGVQPNPVP 420
QY 421 SPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETAAND 480
Db 421 SPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETAAND 480
QY 481 LISPPSGAPLISALDGLGFTHADVTAVPSTIFTSVPVFTTGTSDYAEINPSIIVRAGSF 540
Db 481 LISPPSGAPLISALDGLGFTHADVTAVPSTIFTSVPVFTTGTSDYAEINPSIIVRAGSF 540
QY 541 DPSSQPNDRHVAFTSDGKNWFGQSEPGVTTGGTVAASADGSRFVWAPGDPQPVVYAV 600
Db 541 DPSSQPNDRHVAFTSDGKNWFGQSEPGVTTGGTVAASADGSRFVWAPGDPQPVVYAV 600
QY 601 GFNGSMAASQGVPAANAQIRSDRNPNTFYALNGTFTYRSTDGGVTFQVAAGLPSSGAVG 660
Db 601 GFNGSMAASQGVPAANAQIRSDRNPNTFYALNGTFTYRSTDGGVTFQVAAGLPSSGAVG 660
QY 721 GTIGGVTGAYRSDDCGTTWVINDQHOYGNWQAITGDHANLRVYIGTNGRGIVYCDI 780
Db 721 GTIGGVTGAYRSDDCGTTWVINDQHOYGNWQAITGDHANLRVYIGTNGRGIVYCDI 780
QY 781 GGAPSGSPSVSPSASPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840
Db 781 GGAPSGSPSVSPSASPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840
QY 841 SASPSSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 900
Db 841 SASPSSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 900
QY 901 SSVDLSTVTVRYWTFTRDGGSTLVYNCWAAIGCGNIRASFGSNVNPATPTADTYIQ 956
Db 901 SSVDLSTVTVRYWTFTRDGGSTLVYNCWAAIGCGNIRASFGSNVNPATPTADTYIQ 956
```

RESULT 2

```
US-10-155-400-1
; Sequence 1, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
```

```
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulosyticus
; NAME/KEY: MOD_RES
; FEATURE:
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
; US-10-155-400-1

Query Match 100.0%; Score 5134; DB 14; Length 957;
Best Local Similarity 100.0%; Pred. No. 3.9e-280;
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRSENIRLTWRSRRLVSLAATASFAVAALGLVLPITATSPAHAAATTPYTWSNVAIG 60
Db 1 MDRSENIRLTWRSRRLVSLAATASFAVAALGLVLPITATSPAHAAATTPYTWSNVAIG 60
QY 61 GGGFVDGIVFNEGAGILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 120
Db 61 GGGFVDGIVFNEGAGILYVRTDIGMYRWDAANGRWIPLLDWYGNWNGYGVVIAAD 120
QY 121 PINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLGNNMGRGERLAVD 180
Db 121 PINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLGNNMGRGERLAVD 180
QY 181 PNNDNILYFGAPSGKGLWRSDGATWSQMTNFPDVGTYIANPTDTTGYQSDIOGVVWA 240
Db 181 PNNDNILYFGAPSGKGLWRSDGATWSQMTNFPDVGTYIANPTDTTGYQSDIOGVVWA 240
QY 241 FDKSSSSILGQASKTIFVGVADPNPNVFWSRDGGATWQAVPGAPTGFIPHKGVFDPVNVHL 300
Db 241 FDKSSSSILGQASKTIFVGVADPNPNVFWSRDGGATWQAVPGAPTGFIPHKGVFDPVNVHL 300
QY 301 YIATSNTPGPGYDSSGDVWKFVSVTGWTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360
Db 301 YIATSNTPGPGYDSSGDVWKFVSVTGWTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360
QY 361 VATQISWMPDITIIFRSTDGGATWTRIDWNTSYPNRSRYVLDISAEPLWTFGVQPNPVP 420
Db 361 VATQISWMPDITIIFRSTDGGATWTRIDWNTSYPNRSRYVLDISAEPLWTFGVQPNPVP 420
QY 421 SPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETAAND 480
Db 421 SPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETAAND 480
QY 481 LISPPSGAPLISALDGLGFTHADVTAVPSTIFTSVPVFTTGTSDYAEINPSIIVRAGSF 540
Db 481 LISPPSGAPLISALDGLGFTHADVTAVPSTIFTSVPVFTTGTSDYAEINPSIIVRAGSF 540
QY 541 DPSSQPNDRHVAFTSDGKNWFGQSEPGVTTGGTVAASADGSRFVWAPGDPQPVVYAV 600
Db 541 DPSSQPNDRHVAFTSDGKNWFGQSEPGVTTGGTVAASADGSRFVWAPGDPQPVVYAV 600
QY 601 GFNGSMAASQGVPAANAQIRSDRNPNTFYALNGTFTYRSTDGGVTFQVAAGLPSSGAVG 660
Db 601 GFNGSMAASQGVPAANAQIRSDRNPNTFYALNGTFTYRSTDGGVTFQVAAGLPSSGAVG 660
QY 661 VMFHAVPGKEGDLWLAASGLYHSTNGGSSWNAITGVSSAVNMGFKSAPGSSYPVAVV 720
Db 661 VMFHAVPGKEGDLWLAASGLYHSTNGGSSWNAITGVSSAVNMGFKSAPGSSYPVAVV 720
QY 721 GTIGGVTGAYRSDDCGTTWVINDQHOYGNWQAITGDHANLRVYIGTNGRGIVYCDI 780
Db 721 GTIGGVTGAYRSDDCGTTWVINDQHOYGNWQAITGDHANLRVYIGTNGRGIVYCDI 780
QY 781 GGAPSGSPSVSPSASPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840
Db 781 GGAPSGSPSVSPSASPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840
```


Db 421 APWKGLSEETAANDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTTGSVDY 480
QY 527 AELNPSIIIVRAGSFPSSQPNDRHVAFTSDGGKWFQSGSEPGVTTGTTVAASADGSRFV 586
Db 481 AELNPSIIIVRAGSFPSSQPNDRHVAFTSDGGKWFQSGSEPGVTTGTTVAASADGSRFV 540
QY 587 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGGVTF 646
Db 541 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGGVTF 600
QY 647 QPVAAGLPSSGAVGMFHAVPKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706
Db 601 QPVAAGLPSSGAVGMFHAVPKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
QY 707 KSAPGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWGQAITGDHANLRRV 766
Db 661 KSAPGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWGQAITGDHANLRRV 720
QY 767 YIGTNGRGIVYGDIGGAPSG 786
Db 721 YIGTNGRGIVYGDIGGAPSG 740
RESULT 5
US-10-155-400-3
; Sequence 3, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Catalytic domain GH74
US-10-155-400-3
Query Match 78.6%; Score 4036; DB 14; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.5e-218;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVVRTDIGMYRWDAAANGRWIPLLDWVGW 106
Db 1 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVVRTDIGMYRWDAAANGRWIPLLDWVGW 60
QY 107 NNWNGYGVVSIADPINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLG 166
Db 61 NNWNGYGVVSIADPINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLG 120
QY 167 NMPGGMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 226
Db 121 NMPGGMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180
QY 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 286
Db 241 IPHKGVFDPVNHVLIATSNTPGPDGSSGDVWKFVSTSGTWTRISPVSTDTANDYFCY 300
QY 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGGATWTRINDWTSYPNRSRLRYVLDSAE 406

Db 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGGATWTRINDWTSYPNRSRLRYVLDSAE 360
QY 407 PMLTFGVQPNPVPSPKLGWMDMAIDPENSRLMLYGTGATLYATNDLTKWDSGGQIHI 466
Db 361 PMLTFGVQPNPVPSPKLGWMDMAIDPENSRLMLYGTGATLYATNDLTKWDSGGQIHI 420
QY 467 APWKGLSEETAANDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTTGSVDY 526
Db 421 APWKGLSEETAANDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTTGSVDY 480
QY 527 AELNPSIIIVRAGSFPSSQPNDRHVAFTSDGGKWFQSGSEPGVTTGTTVAASADGSRFV 586
Db 481 AELNPSIIIVRAGSFPSSQPNDRHVAFTSDGGKWFQSGSEPGVTTGTTVAASADGSRFV 540
QY 587 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGGVTF 646
Db 541 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIRSDRVNPKTFYALNSGTFFYRSTDDGGVTF 600
QY 647 QPVAAGLPSSGAVGMFHAVPKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706
Db 601 QPVAAGLPSSGAVGMFHAVPKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
QY 707 KSAPGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWGQAITGDHANLRRV 766
Db 661 KSAPGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWGQAITGDHANLRRV 720
QY 767 YIGTNGRGIVYGDIGGAPSG 786
Db 721 YIGTNGRGIVYGDIGGAPSG 740
RESULT 6
US-10-155-400-6
; Sequence 6, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; OTHER INFORMATION: Catalytic domain GH74
US-10-155-400-6
Query Match 78.6%; Score 4036; DB 14; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.5e-218;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVVRTDIGMYRWDAAANGRWIPLLDWVGW 106
Db 1 ATTQPYTWSNVAIGGGGFVDGIVFNEGAPGILYVVRTDIGMYRWDAAANGRWIPLLDWVGW 60
QY 107 NNWNGYGVVSIADPINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLG 166
Db 61 NNWNGYGVVSIADPINTNKVAAVGMVNTNSWDNDGAILRSSDQATWQITPLPFKLG 120
QY 167 NMPGGMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 226
Db 121 NMPGGMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180
QY 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 286
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 240

QY 287 IPHKGVEPDPVNHVLIATISNTGGPYDSSGDMWKFESVTSCTWTIRISPVSTDTANDYEGY 346
DB 241 IPHKGVEPDPVNHVLIATISNTGGPYDSSGDMWKFESVTSCTWTIRISPVSTDTANDYEGY 300
QY 347 SGLIIRDRQHPNTIMVATQISWMPDTIIPRSTGCGATWTRIMWDTSYNRSIRYVLDISAE 406
DB 301 SGLIIRDRQHPNTIMVATQISWMPDTIIPRSTGCGATWTRIMWDTSYNRSIRYVLDISAE 360
QY 407 PWTFTGQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466
DB 361 PWTFTGQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
QY 467 APWKGLBETAVNDLISPPSGAPLISALDGGTHADVTAVPSTIPTSPTFTGTSVDY 526
DB 421 APWKGLBETAVNDLISPPSGAPLISALDGGTHADVTAVPSTIPTSPTFTGTSVDY 480
QY 527 AELNPSIIVRAGSFPDSSQPNDRHVARSTGCKNWFQSGEPGVTTGGTVAAASADGSRFV 586
DB 481 AELNPSIIVRAGSFPDSSQPNDRHVARSTGCKNWFQSGEPGVTTGGTVAAASADGSRFV 540
QY 587 WAPDGPQPVVYVGVFGNSWAASQGVANAOIRSDRVNPKTFYALNSGTFFYRSTGCVTF 646
DB 541 WAPDGPQPVVYVGVFGNSWAASQGVANAOIRSDRVNPKTFYALNSGTFFYRSTGCVTF 600
QY 647 QPVAAGLPSSGAGVGMFHAIVPGKBDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706
DB 601 QPVAAGLPSSGAGVGMFHAIVPGKBDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
QY 707 KSAPOSSYPVAVVGTIGVTCAYRSDCGTTWLLINDDQHOYGNWQAITGDHANLRRV 766
DB 661 KSAPOSSYPVAVVGTIGVTCAYRSDCGTTWLLINDDQHOYGNWQAITGDHANLRRV 720
QY 767 YIGTNGRGIVYGDIGGAPSG 786
DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 7
US-10-156-761-9395
; Sequence 9395, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9395
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9395
Query Match 48.3%; Score 2478; DB 14; Length 882;
Best Local Similarity 50.4%; Pred. No. 7e-131;
Matches 473; Conservative 128; Mismatches 227; Indels 110; Gaps 14;
QY 23 TASFAVAAL--GVLPATATAPAHAAATQPTWNSVAGGGFVDGIVFNEGACILYV 80
DB 8 TAVLALAGLPATPALAASAPTATTAADTYSWKNAVVDGGFVPGIVFNRSEKNLAYA 67

QY 81 RTDIGMYRWAANGRWIPLLDWGMNNGVNGVVSIAADPINTNKVAAVGMVYTNWD 140
DB 68 RTDIGMYRWAANGRWIPLLDWGMNNGVNGVVSIAADPINTNKVAAVGMVYTNWD 127
QY 141 NDGAILRSDGATWQITPLPFLKGNMPCRMGERLAVDPNNDNLLYFGAPSGKGLWRS 200
DB 128 GNGAVLRSGDRGASQKTDLPFLKGNMPCRMGERLAVDPNNDNLLYFGAPSGKGLWRS 187
QY 201 TDSGATWQMTNPPDVGTYIANPTDTTGYOSDIQGVVWVAFDKSSSSSLQOAKSTIFVGV 260
DB 188 TDSGASWQVTDVFPNVGTYVQATDTSGYASNDQGLVWTFDEBTSFGSSFTFVYVGV 247
QY 261 DPNNFVFRSGGATQWAVPGAPTGFIPHKGVFDPVNHVLYIATNTGPGYDSSGDMVK 320
DB 248 DXONSIVYSTDAGATWSRLAGQPTGHLAHKGVLDAAANGCLYLAYSCKGPGYDGGKQLWR 307
QY 321 FSVTSGTWTRISPVSTDTANDYFGYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTG 380
DB 308 YTTKTGTWNTISPVAAEDT--YGFSGLTVDROHPGTVMATAYSSWMPDTQLFRSTDSG 364
QY 381 ATWTRIMWTSYPNRSLRVLDISAEPMLTFCVQPNPVPSPKLGWMDMAIDPFNSDR 440
DB 365 GTWTKANDTYSYPSNRFTMDVSSPMLTGANPAPPEQTCKLGWMTSLEIDPFDSAR 424
QY 441 MLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDLISPPS--GAPLISALDGLG 499
DB 425 MMYTGAVYGTNDLNTWDSGQFTIKPNARGLEETAVNDLASPPSGGQALFSAIDGIG 484
QY 500 FTHADVAVPSTIFTSPVFTTCTSDYAEINPISIIVRAGSFPDSSQPNDRHVAFASTDDGK 559
DB 485 FRHTDLTTPVPSLMYSPNFTTSTSLDYAETDPCTVVRVGNLD--SGP--HVAFASTDNGA 539
QY 560 NWFQSGEPGVTTGGTVAAASADGSRFVWAPGQPVVYVAVFGNSWAASQGVANAOIR 619
DB 540 NWFAGADPSGSGGTVAAASDGRFVWSPAGTG--VQYTTGFTGTSWASAGLPAGAIVE 597
QY 620 SDRVNPKTFFYALNSGTFFYRSTGCVTFQPVAA--GLPSSGAGVGMFHAIVPGKBDLWLAAS 678
DB 598 SDRVDPKTFYFGKGRFVYSSDGGATFTASAATGLFSGDS--VRFKALPCKGDIWL 655
QY 679 S-----GLYHSTNGSSWSAITGVSSAVNVGFGKSAFSGSYPAVFFVGTIGGVTGVRSD 733
DB 656 ASDGAYGLWHSTDDGGAFTKLATVQDADTIGFKAATGASYQTLTYSAKIGGVRGIFRST 715
QY 734 DCGTTWLLINDDQHOYGNWQAITGDHANLRRVYIGTNGRGIVYGDIGGAPSGSPSPSV 793
DB 716 DKGASWTRVNDDAHQWGTGAATIGDPRVYVYVSTNGRGIVYGDITAGSSDGG----- 769
QY 794 PSASPLSP 853
DB 770 ----- 769
QY 854 SSPSSSP 913
DB 770 ---GTEPAT-----GACTVYIRITNWSGFGQ--ADVQLANTGSTANDGWSLG-- 814
QY 914 FTRDGSSTLVYNCDMAAIGCG-----NIRASFGS 943
DB 815 SFGDGQEVTLWNASVYQAAGSGVTAANLAWNGRVAAGS 852
RESULT 8
US-09-917-376-7
; Sequence 7, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOGLYTICUS
; FILE REFERENCE: 40197.40501


```
RESULT 10
US-10-420-191-2
; Sequence 2, Application US/10420191
; Publication No. US20040067569A1
; GENERAL INFORMATION:
; APPLICANT: Haas, Jeffrey A.
; APPLICANT: Rev. Michael W.
; APPLICANT: Zatecky, Elizabeth J.
; APPLICANT: Haas, Jeffrey A.
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids
; TITLE OF INVENTION: Encoding Same
; FILE REFERENCE: 10210.200-US
; CURRENT APPLICATION NUMBER: US/10/420,191
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: US 60/373,987
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-420-191-2

Query Match 31.7%; Score 1625.5; DB 15; Length 838;
Best Local Similarity 41.7%; Pred. No. 5.5e-83;
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;

QY 25 SPFAVAALGLPIALTAGPAHAATQPYTWSNVAI-GGGGFDGIVFNEGAPGILYVRTD 83
DB 4 SRVLALVGV-----IPAHAA-----PSWKVNLGGGGFVPGIIFPKTKGVAYARTD 53

QY 84 IGMVTRWDAANGRWIPLLDWV-----GMNNGVGVVSIADPINTNKVAAVGMVNTNSWD 139
DB 54 IGLYLRLN-ADDSWTAVTGDIADNAGWNNW---GIDAVALDQDDQKVAAVGMVNTNSWD 109

QY 140 PNDGAILRSSDGGATWQITPLPKLGGNMPGMRGERLAVDNNDNLLYFGAPSGKGLWR 199
DB 110 PSNGAIIRSSDRGATWSFTNLFPKVGNNMPGMRGERLAVDNNDNLLYFGAPSGKGLWK 169

QY 200 STDSGATWSOMTNFDPVGYIANPTDTTGYQSDIQGVVWVAIPKSSSSLSQASKTIFVGV 259
DB 170 STDGGVTSKVSFTATGYIIPDSDNGYNDKQGLMWVTFDSTSTTGATSRIFVGT 229

QY 260 ADP-NNPVWSDGGATWQAVPGAPTGFIPKGVFDPVNVHLYIATSNTPGPDGSSGDV 318
DB 230 ADNITASVVVSTNAGSTWSAVPGQPKYFPKAKLQPAEKALYLYTSDGTGPDGTLGV 289

QY 319 WFSVTSQGTWTRISVPSTDTANDYFGYSGGLIDROHPNTIIVWATQISWPDITIPRSTD 378
DB 290 WRYDIAGGTWKDITFVSGSDL---YFGGGLGLDLQKPGTLVVASLNSWPPAQLPRSTD 346

QY 379 GGNATWTRWDTSYNRSRLRYVLDISAEPLTFG-VQPNPPVPS---PKLGMWDEAMAI 433
DB 347 SGTWSPIMAWASYPETETYYISITPKAPWIKNNFIDVTSESPDGLIIRKLGMMLESLEI 406

QY 434 DPNRSRLMYGTGATLYATNDLTKWDSGGQIHIAPMVKGLIETAVNDLISPPSGAPLISA 493
DB 407 DPTDSNHWLYGTGMIIFGCHDLTNWDTNRHNSIQSLADGIEFVSQDLASAPGSELLAA 466

QY 494 LGDLGGFTHA---DVTAVPSTIFTSPVFTTGSVDVAELNPSIIVRAGSDFDSSQPNDRH 550
DB 467 VGDNGFTFASRNDLGTSPQVWATPTWATSTSDVYAGNSVKSVRVGNVGTAGTQQ----- 521

QY 551 VAFSTDGGKWNQGSPEPGVTTGGTVAASADSRFWAPCDQPPVYAVGFGNSWAASQ 610
DB 522 VAISDGGATWSDIDYAADTSMNGGTVAISADGDTILWSTASSG---VQRSQFGSGFASVS 578

QY 611 GVPANAQIRSDRVNPKTFYALNGTFRYSTDGGVTFQPVAAAGLPPSSGAVGVW---FHAVPG 668
DB 579 SLPAGAVIASDKTNSVVFAGSGSTFVYVKDGTSSP---TRG-PKLGSGATIRDIAAHT 634

QY 669 KEGDLMAASSGLYHSTNGGSSWSAI-TGVSSAVNVVGFSGKAPGSSYPVAVFVGTGGVTT 727
```

```
DB 635 TAGTLYVSTDVGIIFRSTDSGITTFGQVSTALTINTYQIALGVGS-GSNW-NDYAFGT--GPS 690
QY 728 GA--YRSDDCGTTWVLINDDOHQYGNWQAIITGDHANLRRVYIGTNGRGIVY--GDICGA 783
DB 691 GARLYASGDSGASWTDIQSQGFGSIDSTKVAGSGTAGQVYVGTNGRGVYAGTGVGG 750
QY 784 PSGSPSPVSPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 843
DB 751 TGGT-----SSSTKQSSSTSSASSSTTLRSSVWSTTRASTVTSRTSSAAGTGS 801

RESULT 11
US-10-026-994-2
; Sequence 2, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: ECVI Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-026-994-2

Query Match 31.2%; Score 1604; DB 14; Length 818;
Best Local Similarity 41.8%; Pred. No. 8.6e-82;
Matches 340; Conservative 128; Mismatches 291; Indels 54; Gaps 21;

QY 52 YTWSNVAI-GGGGFDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWV---GW 106
DB 2 FSWKNVNLGGGGFVPGIIFPKTKGVAYARTDIGLYRLN-ADDSWTAVTGDIADNAGW 60

QY 107 NMNGYGVVSIADPINTNKVAAVGMVNTNSWPNDCGAILRSSDGGATWQITPLPKLGG 166
DB 61 HNW---GIDAVALDQDDQKVAAVGMVNTNSWPNDCGAILRSSDGGATWQITPLPKLGG 117

QY 167 NMFGRMGERLAVDNNDNLLYFGAPSGKGLWSTDSGATWSOMTNFDPVGYIANPTDT 226
DB 118 NMFGRMGERLAVDNPNANIIYFGARSGNGLWKSITDGGVTFKVSSTFATGTGYIPDPSDS 177

QY 227 TGYOSDIQGVVWVAIPKSSSSLSQASKTIFVGVADP-NNPVWSDGGATWQAVPGAPT 285
DB 178 NGYNSDKQGLMWVTFDSTSTTGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPK 237

QY 286 FIPKGVFDPVNVHLYIATSNTPGPDGSSGDVWKFVTSQGTWTRISVPSTDTANDYFG 345
DB 238 YFHKAKLQPAEKALYLYTSDGTGPDGTLGVSVRYDIAGGTWKDITFVSGSDL---YFG 294

QY 346 YSLGTIDROHPNTIIVWATQISWPDITIPRSTDCGATWTRIDWTSYPNRSLRYVLDISA 405
DB 295 FGLGLDLQKPGTLVVASLNSWPPAQLPRSTDGTTSPSPIMAWASYPETETYYISITPK 354

QY 406 EPLMTFG-VQPNPPVPS-----PKLGMWDEAMAIIDPNSSDRMLYGTGATLYATNDLTKWDS 460
DB 355 APWIKNNFIDVTSESPDGLIIRKLGMMLESLEIDPTDSNHWLYGTGMIIFGCHDLTNWDT 414

QY 461 GGGIHIAPMVKGLIETAVNDLISPPSGAPLISALGDLGGFTHA---DVTAVPSTIFTSPV 517
DB 415 RHNVS:QSLADGIEFVSQDLASAPGSELLAAVGGDNGFTFASRNDLGTSPQVWATPT 474

QY 518 FTTGTSVDVAELNPSIIVRAGSDFDSSQPNDRHVAFTSDGKKNWFQSGEPGVTGGTVA 577
DB 475 WATSTSDVYAGNSVKSVRVGN-----TAGTQVAISSDGGATWSDIDYAADTSMNGTVA 528
```


QY 578 ASADGSRFVWAPGDPQPVVYAVFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFY 637
 DB 529 YSADGDTILNSTASSG---VORSQFQGSFASVSLPAGAVTASDKTNSFYAGSGSTFY 585
 QY 638 RSTDGGVTFQPVAAAGLPSGAVGM--FHAVPGKEGDLWLAASSGLYHSTNGSGSWSAI- 694
 DB 586 VSKDTGSSF--TRG-PKLSAGTIRIDIAHPTTAGTLYVSTDVGIIFRSTDSGTTFGQVS 641
 QY 695 TGVSAAVNGFGKAPSSPVPVAVFVGTIGVTCG--YRSDDCGTTWVLINDDOHQYGNW 752
 DB 642 TALNTNYQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASMTDIQSGQGFSGID 697
 QY 753 GOALTGPHANLRRVYVINGNKGIVY--GDICGAPSGSPSPVSPSASPSLSPSPSPSSP 810
 DB 698 STKAGSGTAGQVYVINGNKGIVYAGTGGGTGT-----SSSTKQSSSTSSA 748
 QY 811 SPSPSPSSPSSPSPSPSPSPSPSPSPSPSPSAS 843
 DB 749 SSSTTLRSSVVTTRASTVTSRTSSAAGPTGS 781

RESULT 12

US-10-156-761-10111
 ; Sequence 10111, Application US/10156761
 ; Publication No. US20030119018A1

GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 10111
 ; LENGTH: 739
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-10111

Query Match 28.1%; Score 1442; DB 14; Length 739;
 Best Local Similarity 40.8%; Pred. No. 9.7e-73;
 Matches 328; Conservative 107; Mismatches 272; Indels 96; Gaps 25;

QY 7 IRLTMRRLVSLAATASPAVAALGVLPITAITASPAHAATT---QPYTWSNVAIGGG 63
 DB 1 MTRPRRTV--LAGTAA---AALATVP--AVGSAHAETTAGPSYRWRNAVIGTG 53
 QY 64 FVDGIVFNEGAPGILYVTRDTIGMYRDAANGRWIPLDWMYNNWYGVVSTAADPIN 123
 DB 54 FVTGVLFHPSVRGLAYARTDIGGAYRWDGRTPLDLHLGWDNLLGVEAMAVDPTH 113
 QY 124 TNKVAAGVMTNSWDNDGAILRSSDQATWQITPLPFKLGNNMPGMRGLAVDPNN 183
 DB 114 PDRLYLAVGTYAQSAGN-GAVLRSEDGRTWTRTDLTVKLGNGEDGGRGGERLLVDP 172
 QY 184 DNILYFGAPSKGLWRSTDGATSQMTNFDVGTYIANPTDTTGYQSDIQGVVWAFDX 243
 DB 173 SDTLWLGT-RHUGLTKSTDGATWAAATAFP-----AKANSSGQGVF----- 214
 QY 244 SSSSILGQASKTIFVGVADPNPNP-----VWGRDGGATWQAVPGAPTGF---IPHKGVDP 295
 DB 215 ----LVAAGRTVYAGWGDDGTSGTANLYRTAD-GTTWGAVGRPSGTSAKVPLRAAYDT 269

QY 296 VNHVLYIATSNNTGGPYDSSGDVWKFVSTSGTWTTRISPV-----PSTDTANDVFGYSGLTI 351
 DB 270 HTRELYTVYTGADPGPGGSGDGVHKLRTATGTWTEVTVPKPGTTS DGSADTFAYGGVAV 329
 QY 352 DRQHPNTIMVATQISWNPDTIIFRSTDGGAATWTRIMDWTSYNSRSLRYVLDISAEPWLTF 411
 DB 330 DARRPGLTLVSTNNRWADGDTVFRSTDGRTWTSKLD-----AAVFDVSETPFLDW 380
 QY 412 GVQPNPPVPSPKLGWMDMAIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVK 471
 DB 391 GDD-----KPAFGHWIQALAVDPDSQHVYVGTGATLYGTRDLKRW-----APRIR 426
 QY 472 GLEETAVNDLISPPSG-APLISALGDLGCFTHADVTAVPST-IFTSVPVFTTGTSDVYAE 529
 DB 427 GLEESAVRQLISPPVGEAHLISGLDIGVMYHERLTASPSRGMATNPFVFSATGLAQAAA 486
 QY 530 NPSIIVRAGSPDPSSQPNDRHVAFTDGGKMW--PQSEPGGVTTTGGTVAASADGSRFVW 587
 DB 487 RPAYVVRTGWGDHNG-----AYSHDGGRTWAPFEAQPIAKADAPGIATSDGGLTW 540
 QY 588 A-----PGDPGQPVVA---VGFNGNSWAASQGVPAQAQIRSDRVNPKTFVA--LSNGTFY 637
 DB 541 SFVHWG-----TTAAHRSTNDGASWSEVSSFPKGATPVADPADTRFYAYDFDNGTLY 595
 QY 638 RSTDGGVTFQPVAAGLPSSGAVGMFHAVPGKEGDLWLAAS-SGLYHSTNGSGSWSAITG 696
 DB 596 ASTDSGRSFTARAGGLP-SGDSQFKLVAAPGRSGDLWLSAKWNGLYRSTDDGDTFARIDS 654
 QY 697 VSSAVNVGFGKAPGSSYPVAVFVGTIGVTCAYRSDDCGTTWVLINDDOHQYGNWQAI 756
 DB 655 CWASYTLGLGKAADGADYPAIYQVGSTETITAYRSDDAARTWVRINDDAHGWGIGEAV 714
 QY 757 TGDHANLRRVYVINGNKGIVYGD 779
 DB 715 VGDPRHGRVYLATNGRGIQYGE 737

RESULT 13

US-10-395-241-12
 ; Sequence 12, Application US/10395241
 ; Publication No. US20040038367A1
 ; GENERAL INFORMATION:

APPLICANT: MITSUISHI, Yasushi
 ; APPLICANT: YAOI, Katsuro
 ; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
 ; FILE REFERENCE: Q73756
 ; CURRENT APPLICATION NUMBER: US/10/395,241
 ; CURRENT FILING DATE: 2003-03-25
 ; PRIOR APPLICATION NUMBER: JP 2002-83433
 ; PRIOR FILING DATE: 2002-03-25
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 12
 ; LENGTH: 812
 ; TYPE: PRT
 ; ORGANISM: Geotrichum sp. M128
 US-10-395-241-12

Query Match 21.9%; Score 1125; DB 15; Length 812;
 Best Local Similarity 34.5%; Pred. No. 7.2e-55;
 Matches 285; Conservative 129; Mismatches 291; Indels 120; Gaps 31;

QY 31 ALGVLPITAITASPAHAATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRDTIGMYRW 90
 DB 6 SLGKALTALSILASLAVAKEHYEFKNVAIGGGYITGIVAHPKTKDLYARTDIGGAYRW 65
 QY 91 DAANGRWIPLDWMYNNWYGVVSIADPINTNKVWAAVGMYT-NSWDPNDGAILRSS 149
 DB 66 DAGTSKWITPLNDFIEAQDMNINMGTSIALDPNPDRLYLAQQRYVGDEW----AAFVSE 121
 QY 150 DQATWQITPLPFKLGNNMPGMRGLAVDPNDNDILYFGAPSGKGLWRSTDSGATWSQ 209

Db 122 DRGQSFTIYESPPFPMGANDMGRNGLAVNPNFNSNEVMGTRT-EGIKWSSDRAKTWTN 180
Qy 210 MTFNPVGVTYIANPTDTTGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWS 269
Db 181 VTSIPDAF-----TNGIGYTS-----VIFDP-----ERNGTIYASATAPOG-MYVT 220
Qy 270 RDGGATQAVPGAPTGT-----PHKGVFDPVNVHLYIATSNITGPGY 311
Db 221 HDGVSWEFVAGQSSWLNRTTGAFFDKKPSIAPOPMKVALT-----NELYVYIADYPPGPW 278
Qy 312 DGSSGDVWKFVSTGTWTRI-----SPVSTDTANDYFYSGLTIDROHPNTIMVATQ 364
Db 279 GVTFCVWRQNTSGAWDDITPRVGNSSPAPVNNQTPPAGFCGLSVDATNPNRLVWIT- 337
Qy 365 ISWMPDTI-----IFRSTDCGATWTRIDWTS-----YPNRSLRYVLDISAEWLPFG 412
Db 338 LDRDPGALDSIYLTSDAGATKQVTLSPSPNLEGNWGHPTNAARY-KDGTVPVWLPDN 396
Qy 413 VQP-----NPPVPS-----KLGMDEAMALDPFNSDRMLYGTGATLYATNDLTK-----WDS 460
Db 397 NGPQWGGYGAPHGTPGLTKFGWMSAVLIDPFNPEHLMYGTGATWATDLSRVEKDW-- 454
Qy 461 GQOIHIAP-----MVKGLEBETAVNDLISPPSGAPLISALDGLGFTHADVTAVPSTFTSP 516
Db 455 -----APSWYLIQIDGIEENALSLRSPKSGAALLSGIGDISGMKHDLLTK-PQRMFGAP 507
Qy 517 VFTTGSVDYAEINPSIIVRAGSPDPQNDRHVAFSTDDGKNW---FQSGEPGGVTT--- 572
Db 508 QPSNLDSDAAGNFNPNVVRAGSSGHEYDSACARGAYATDGDGDAWTFPTCPQGNWASHY 567
Qy 573 -GGTVAASADGSRFVWAP--GDPGPVYVAVFGNSWAAASQGVPA-----NAQIRSDRV 623
Db 568 QGSTIADVASGSIWSTKLDEQASGPWYSHDYGKTS-----VPAGDLKAQNTANVLSDKV 623
Qy 624 NPKTYALSNGTFRSTDDGVTQFVAAAGLPSGAGVGMFHAVPGKEGDLWL-AASSGLY 682
Db 624 QDGTFTYADGKFFVSTDDGKSYAAKAGLVLT--GTSMLPVPVWAGDVVVPVEGGLP 681
Qy 683 HSTNGSSWSAI-TGVSSAVNVFGKS-----APGSSYPVAVFVGT--IGGVTGAYRSDDC 735
Db 682 HSTDRGASFTRVGTANATLVSVGAPKSDGKASAPSAVFITWGTDPKGSIDGLYRSDN 741
Qy 736 GTTWVLINDDQHOYGNWQGAITGDHANLRVYIGTNGRGIVYDI 780
Db 742 GSTWTRVNDQEHNYSG-PTMIEADPKYGRVYLTNGRGIVYADL 785

RESULT 14
US-10-395-241-14
; Sequence 14, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Geotrichum sp. M128
US-10-395-241-14

Query Match 21.8%; Score 1117; DB 15; Length 789;
Best Local Similarity 35.0%; Pred. No. 28-54; Indels 120; Gaps 31;
Matches 281; Conservative 124; Mismatches 279;

Qy 52 YTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWVGWNNWGY 111
Db 4 YEFKNAIGGGYITGIVAHPTKOLLYARTDIGGAYRWDACTSKWIPLDFTFEADQMN 63
Qy 112 NGVSIADPINTNKVWAAVGHMT-NSWDPNDGAILRSSDOGATWQITPLPKLGGNMPG 170
Db 64 MGTESIALDPNPNDRLLYLAQGRYVGDEW---AAFVSEDRGQSFTIYESPPFPMGANDMG 119
Qy 171 RGMGERLAVDPNNDILYFGAPSGKGLMRSTDSGATWSOMTNFPDVGTYIANPTDTTGYQ 230
Db 120 RNNGERLAVPNPNSNEVMGTRT-EGIKWSSDRAKTWTNVTIPDAF-----TNGIGYT 172
Qy 231 SDIQGVWVWAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGTI- 287
Db 173 S-----VIFDP-----ERNGTIYASATAPOG-MYVTHDGGVSEPVAGQSSWLNRT 218
Qy 288 -----PHKGVFDPVNVHLYIATSNITGPGYDGSGLDGVWKSFSVSTGTWTRI- 331
Db 219 TGAPFDKPPASTAPOPMKVALT-----NELYVYIADYPPGPWGTGVEYWRQNTSGAWDDIT 276
Qy 332 -----SPVSTDTANDYFYSGLTIDROHPNTIMVATQISWMPDTI-----IFRSTDCGAT 382
Db 277 PRVGNSSPAPVNNQTPPAGFCGLSVDATNPNRLVWIT-LDRDPGALDSIYLTSDAGAT 335
Qy 383 WTRIWDWTS-----YPNRSLRYVLDISAEWLPITFGVQP-----NPPVPS-----KL 425
Db 336 WKDVTQLSSPNSLEGNWGHPTNAARY-KDGTVPVWLPDNFNGPQWGGYGAPHGTPGLTKFG 394
Qy 426 WMDEAMALDPFNSDRMLYGTGATLYATNDLTK-----WDSGGQIHIAP-----MVKGLEBETA 477
Db 395 WMSAVLIDPFNPEHLMYGTGATWATDLSRVEKDW-----APSWYLIQIDGIEENA 446
Qy 478 VNDLISPPSGAPLISALDGLGFTHADVTAVPSTFTSPVFTTGTSDVDAELNPSIIVRA 537
Db 447 ILSLRSPKSGAALLSGIGDISGMKHDLLTK-PQRMFGAPQFNSLDSIDAAGNFNPNVVR 505
Qy 538 GSFDPSSQPNDRHVAFSTDDGKNW---FQSGEPGGVTT---GGTVAASADGSRFVWAP--G 590
Db 506 GSGGHEYDSACARGAYATDGDGDAWTFPTCPQGNWASHYQGSTIADVASGSIWSTKL 565
Qy 591 DPGQPVYVAVFGNSWAAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFRSTDDG 644
Db 566 EQASGPWYSHDYGKTS-----VPAGDLKAQNTANVLSDKVDGTFYATDGGKFFVSTDDGK 621
Qy 645 TFPQVAAAGLPSGAGVGMFHAVPGKEGDLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN 702
Db 622 SYAAKAGLVLT--GTSMLPVPVWAGDVVVPVEGGLFHSSTDGASFTRVGTANATLV 679
Qy 703 VGFSGS-----APGSSYPVAVFVGT--IGGVTGAYRSDDCGTTWVLINDDQHOYGNWQGA 756
Db 680 VGAPKSDGKASAPSAVFITWGTDPKGSIDGLYRSDNNGSTWTRVNDQEHNYSG-PTMI 738
Qy 757 TGDHANLRVYIGTNGRGIVYDI 780
Db 739 EADPKYGRVYLTNGRGIVYADL 762

RESULT 15
US-10-395-241-18
; Sequence 18, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; PRIOR FILING DATE: 2003-03-25
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Geotrichum sp. M128
US-10-395-241-18


```
; SEQ ID NO 18
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Kyloglucan Oligosaccharide-Degrading Enzyme with Histidine Tag
US-10-395-241-18

Query Match      21.7%; Score 1114; DB 15; Length 826;
Best Local Similarity 35.0%; Pred. No. 3e-54;
Matches 281; Conservative 123; Mismatches 280; Indels 120; Gaps 31;

QY 52 YTSNVAIGGGGFDGIVFNEGAGIILYVRTDIGMYRWDAANGRWIPLLDWVGWNNWGY 111
DB 5 YEFKVAIGGGYITGIVHPKTKDILYARTDIGAYRWDAAGTSKWIPLNDPIEAQDMNI 64

QY 112 NGVSIAADPINTKNVAAVGMT-NSWDPNDAILRSSDQAGATWQITPLPKLGGNMPG 170
DB 65 MGTESIALDPNPDRLVLAQRYVGDEW---AAFVYSEDRGQSFTIYESPPFMGANDMG 120

QY 171 RGMERLAVDPNNNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGYQ 230
DB 121 RNNGERLAVNPFNEVNMGTGT-EGIKSSDRAKTWNVTISIPDAF-----TNGIGYT 173

QY 231 SDIQGVVVAFPDKSSSLGQASKTIFVGVAADPNPNPFWMSRDGGATWQAVPGAPTGFI--- 287
DB 174 S-----VIFDP-----ERNGTIYASATAPQG-MVYTHDGGVSWEPVAGQPSWLNRT 219

QY 288 -----PHKGVFDPVNHVLYIATSNITGGPYDGGSSGDVWKFVSITSGTWTRI- 331
DB 220 TGAPFDKKPASIAPOPMKVALTP-NFLYVYADYPGWPGVTFGKVMRQNRNRTSGAWDDIT 277

QY 332 -----SPVSTDTANDYFGYSGLTIDROHNTIMVATQISWPPDTI----IFRSTDDGAT 382
DB 278 PRVGNSSPAPYNNQTFPAGGFCGLSDVATNENRLNVIT-LDRDPGALDSIYLTSDAGAT 336

QY 383 WTRTWDTST-----YPNRSURYVLDISAEPLWTFGVOP-----NPPVPSP---KLG 425
DB 337 WKDVTQLSSPNLEGNWGHPTNAARY-KDGTVPVPLDNNNGPQWGGYGAPHTPGLTKFG 395

QY 426 WMDRAMAIDPFNSDRMLYGTGATLYATNDLTK----WDSGGQIHIAP-----MVKGLEETA 477
DB 396 WMSAVLIDPFPPEHLMYGTGATITWATDLSRVEKDW-----APSWYLQIDGIEENA 447

QY 478 VNDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPTFTGTSVDYAEINPSIIVRA 537
DB 448 ILSLRSPKSGAALLSGIGDISCMKHDDLTK-PQRMFGAPQFSNLDSDAAGNFPNVVRA 506

QY 538 GSPDPSSQPNDRHVAFSTDGCKW--FQSGEPGGVTT---GCTVAASADGSRFYWAP--G 590
DB 507 GSSGHEYDSACARGAYATDGDAMTIFPTCPGNNASHYQGSTIADVAGSQIYVWSTKLD 566

QY 591 DPGQPVVYVAVGFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFFYRSTDGGV 644
DB 567 EQASGPWYSHDYGKTWS-----VPAGDLKAQANTANVLSDKVQDCTFYATDGGKEFFVSTDGK 622

QY 645 TFQPVAAGLPSGAVGVNMFHVPKKEGDLWL-AASSGLYHSTNGGSSWSAI-TGVSSAVN 702
DB 623 SYAAKAGLAVT--GTSLMPAVNPVWAGDVVPVPEGLFHSDFGASFTRVGTANATLVS 680

QY 703 VGFGKS-----APGSSYPAVFVGT--IGGVYGYRSDDCGTTWVLINDDHOYGNWQAI 756
DB 681 VGAPKSKSDGKKASAPSAVFITWGTDKPGSDIGLYRSDDNGSTWTRVNDQEHNSG-PTWI 739

QY 757 TGDHANLRVRYIGTNGRGIVYGD 780
DB 740 EADPKYGRVYLTNGRGIVYADL 763
```

Search completed: October 5, 2005, 08:08:27
Job time : 125.54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 07:09:01 ; Search time 12598.3 Seconds
(without alignments)
3680.789 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENRLTWRSRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09917376/runat_04102005_164328_26897/app.query.fasta_1.2524
-DB=GenEmbl -ORF=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CGN 1 1 15287 @runat_04102005_164328_26897 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl :

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	2869	6	AX700058 Sequence
2	2493	48.5	296500	1	SC0939128
3	2487	48.4	299800	1	AP005028 Streptomy
4	2419.5	47.1	3262	1	AF078038s1

5	2174	42.3	2823	6	AX565635 Sequence
6	2037	39.7	12732	1	AE007608 Clostridi
7	2007.5	39.1	2950	1	AJ585344 Clostridi
8	1784	34.7	2579	8	AJ292929 Agaricus
9	1782.5	34.7	2745	8	AB015511 Aspergill
10	1690	32.9	3959	8	AY040839 Aspergill
11	1633	31.8	2724	8	AK110506 Oryza sat
12	1631.5	31.8	2517	8	AY281371 Hypocrea
13	1570	30.6	299550	1	AP005031 Streptomy
14	1179.5	23.0	14520	1	AE011809 Xanthomon
15	1170.5	22.8	11910	1	AE012276 Xanthomon
16	1161	22.6	13278	1	AE001712 Thermotog
17	1125.5	21.9	2646	6	AX924429 Sequence
18	1125.5	21.9	2716	8	AB089343 Geotrichu
19	1124	21.9	2481	6	AX924435 Sequence
20	1117	21.8	2367	6	AX924431 Sequence
21	993.5	19.3	2755	8	AB116528 Geotrichu
22	823	16.0	3687	8	AX700036 Sequence
23	702	13.7	2289	6	AX700050 Sequence
24	460.5	9.0	590	8	AB1534359 Agaricus
25	431.5	8.4	5437	1	CASR69XN2
26	429	8.4	3365	6	AX700025 Sequence
27	409	8.0	11707	1	AF078737 Caldicell
28	409	8.0	11707	6	E35099 Truncated c
29	407	7.9	299175	1	AP005023 Streptomy
30	402.5	7.8	3237	1	AF200304 Caldibaci
31	398.5	7.8	5513	1	AT286105 A.thermophi
32	397.5	7.7	5439	1	CDCELA Caldocellum
33	391.5	7.6	5284	1	CSU16308 Caldocellum
34	389	7.6	24208	1	AY281357S2 Unculture
35	388.5	7.6	202301	1	AE017286 Desulfocov
36	387	7.5	34615	1	BX569695 Synechoco
37	381.5	7.4	2801	5	AF218784 Gallus ga
38	380	7.4	209091	10	AC122299 Mus muscu
39	372.5	7.3	2015	8	CREVSPJA Chlamydomon
40	370	7.2	206849	2	AC011760 Drosophil
41	369.5	7.2	149159	10	AC148329 Mus muscu
42	368	7.2	242587	10	AC093481 Mus muscu
43	367.5	7.2	215910	10	AC127335 Mus muscu
44	367	7.1	7085	2	AC015260 Drosophil
45	367	7.1	179386	3	AC104515 Drosophil

ALIGNMENTS

RESULT 1	AX700058	Sequence 2 from Patent WO03012090.	2869 bp	DNA	linear	PAT 03-APR-2003
LOCUS	AX700058	Sequence 2 from Patent WO03012090.				
DEFINITION	AX700058					
ACCESSION	AX700058					
VERSION	AX700058.1	GI:29536021				
KEYWORDS	Acidothermus cellulolyticus					
SOURCE	Acidothermus cellulolyticus					
ORGANISM	Acidothermus cellulolyticus					
REFERENCE	1					
AUTHORS	Ding, S. Y., Adney, W. S., Vinzant, T. B. and Himmel, M. E.					
TITLE	Thermal tolerant avicelase from Acidothermus cellulolyticus					
JOURNAL	Patent: WO 03012090-A 2 13-FEB-2003;					
FEATURES	Midwest Research Institute (US)					
source	Location/Qualifiers					
	1. .2869					
	/organism="Acidothermus cellulolyticus"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:28049"					
ORIGIN						
Alignment Scores:						
Pred. No.:	1.14e-146					2869
Score:	5134.00					956
Percent Similarity:	100.00%					0
Best Local Similarity:	100.00%					0

Query Match:	99.98%	Indels:	0
DB:	6	Gaps:	0
US-09-917-376-1 (1-957) x AX700058 (1-2869)			
Qy	1	MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu	20
Db	1	ATGGATCGTTTCGGAGAACATCCGTCGACTATGAGATCAGACGATTTGGTATCACTGCTC	60
Qy	21	AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr	40
Db	61	GCGCCCATCTGCGTGGTTCGCGGTGGCGCGCTCTGGAGTTCTGCCCATCGCATTAACG	120
Qy	41	AlaSerProAlaHisAlaAlaThrThrGlnProIleThrTrpSerAsnValAlaIleGly	60
Db	121	GCITTCCTCTGGGACGCGGCGACGACTCAGCCGTACACTGGAGCAACGTGGCGATCGG	180
Qy	61	GlyGlyGlyPheValAspGlyIleValPheAsnGlyAlaProGlyIleLeuIleVal	80
Db	181	GGCGCGGCTTGTTCGACGGGATCGTCTTCAATGAAGGTGCACCGGGAATTTCTGTACGTG	240
Qy	81	ArgThrAspIleGlyGlyMetTyArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu	100
Db	241	CGACGACATCGGGGGGATGATATCGATGGGATGCGCCCAACGGGCGGTGGATCCCTCTT	300
Qy	101	LeuAspTrpValGlyTrpAsnAsnTrpGlyTyPheAsnGlyValValSerIleAlaAlaAsp	120
Db	301	CTGATTTGGGGGATGGAACAATTTGGGGGTACACCGCGTCGTGAGATTTGGCGACAC	360
Qy	121	ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyTrpAsnSerTrpAspPro	140
Db	361	CCGATCAATACATACAGGTATGGCCCGCTCGGAATGTACACCAACAGCTGGGACCCA	420
Qy	141	AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu	160
Db	421	AACGACGAGCGATTCCTCCGCTCGTCTGATCAGGCGCAACGTGGCAATAACCGCCCTG	480
Qy	161	ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluAtrcLeuAlaValAsp	180
Db	481	CCGTTCAAGCTTGGCGGCAACATGCCCGGGGTGGAAATGGGCGAGCGGCTTGGCGTGGAT	540
Qy	181	ProAsnAsnAspAsnIleLeuTyPheGlyAlaProSerGlyLysGlyLeuTrpArgSer	200
Db	541	CCAAACAATGACAACTTCGTATTTCCGGCCCGCGAGCGGCAAGGGCTCTGGAGAAGC	600
Qy	201	ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyIle	220
Db	601	ACAGATTCGGCGCGACCTGGTCCAGATCGAACTTTCCGACGCTAGGCACGTAACATT	660
Qy	221	AlaAsnProThrAspThrGlyTyTrcGlnSerAspIleGlnGlyValValTrpValAla	240
Db	661	GCAATCCACTGACACACCGCGCTATCAGAGCGATATCAAGGCGCTCGTCTGGGTCGCT	720
Qy	241	PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla	260
Db	721	TTCCGACAAGTCTTCGTATCGCTCGGCAAGCGAGTAGACCAATTTTGTGGGCTGGCG	780
Qy	261	AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro	280
Db	781	GATCCCATATCCGCTCTCTCGAGACAGACGCGCGCGACGCTGGCAGCGGTGGCG	840
Qy	281	GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu	300
Db	841	GGTGGCGGACCGGCTTCATCCGCAACAGGGCGTCTTTGACCCCGTCAACCCAGCTGCTC	900
Qy	301	TyrIleAlaThrSerAsnThrGlyProTyArgAspGlySerSerGlyAspValTrpLys	320
Db	901	TATATTGCCACAGAAATACGGGTGGTCCGTATGACGGGAGCTCCGGGACGCTCTGGAAA	960
Qy	321	PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla	340
Db	961	TTCTCGGTGACCTCCGGGACATGGAGCGGATCAGCCCGGTACCTTCGACGACACGGCC	1020
Qy	341	AsnAspTyPheGlyTySerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet	360
Db	1021	AACGACTACTTTGGTTACAGCGGCTCACTATCGACCGCCAGCACCCGACGATAATG	1080
Qy	361	ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly	380
Db	1081	GTGGCAACCCAGATATCGTGGTGGCGGACACCAATAATCTTTCGGAGCACCGACGGCGT	1140
Qy	381	AlaThrTrpThrArgIleTrpAspTrpTrpSerTyProAsnArgSerLeuArgTyVal	400
Db	1141	CGGAGCTGGACCGGATCTGGGATTTGGAGCTATCCCAATCGAAGCTTGGATATGTG	1200
Qy	401	LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro	420
Db	1201	CTTGACATTTTCGGCGGAGCTTTGGGTGACCTTCGGCGTACAGCCGAATCTCCCGTACCC	1260
Qy	421	SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg	440
Db	1261	AGTCCGAAGCTCGGCTGGATGATGAAGCGATGGCAATCGATCCGTTCAACTCTGATCGG	1320
Qy	441	MetLeuTyGlyThrGlyAlaThrLeuTyAlaThrAsnAspLeuThrLysTrpAspSer	460
Db	1321	ATGCTCTACGGAACAGCGCGAGCTTTGTACGCAACAATGATCTCACGAAGTGGGACTCC	1380
Qy	461	GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp	480
Db	1381	GGCGCCAGATTCATATCGCGCGGATGGTCAAAGGATTTGGAGAGACGGCGTAAACGAT	1440
Qy	481	LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe	500
Db	1441	CTCATCAGCCCGCTCTGGCGCCCGCTCATCAGCGCTCTCGAGACCTTCGGCGGCTTC	1500
Qy	501	ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr	520
Db	1501	ACCCACGCGAGCTTACTGCCGTGCCATCGACGATTTTACGTCAACGGTGTTCACGACC	1560
Qy	521	GlyThrSerValAspTyAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe	540
Db	1561	GGCACCGGCTGACATATCGGAAATTTGAATCCGTGATCATCTTCGCGCTGGAAGTTTC	1620
Qy	541	AspProSerSerGlnProAsnAspArgHisAlaPheSerThrAspGlyGlyLysAsn	560
Db	1621	GATCATTCAGGCAACCCGAACAGCAGCGCTCGCGTCTTCGACAGCGCGCGCAAGAAC	1680
Qy	561	TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla	580
Db	1681	TGGTTCCAGGCGAGCAACCTGGCGGGGTGACAGCGGGCGCACCGTCCGCCGATTCGCC	1740
Qy	581	AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyAlaVal	600
Db	1741	GACGCTCTCGTTTCTGGGCTCCCGGCGATCCCGGTCCAGCTGTGGTGTACGCAATC	1800
Qy	601	GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer	620
Db	1801	GGATTTGGCAACTCTCTGGGCTGTTCCGAAGGTGTTCCCGCAATGCCAGATCCCGTCA	1860
Qy	621	AspArgValAsnProLysThrPheTyAlaLeuSerAsnGlyThrPheTyArgSerThr	640
Db	1861	GACCGGGTGAATCCAAAGACTTTTCTATGCCCTATCCAAATGGAACCTTCTATCGAAGCAG	1920
Qy	641	AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly	660
Db	1921	GACCGCGCGTGACATTCACCGCTCGCGCGCGGTCTTCCGAGCAGCGGTGCGTCCGT	1980
Qy	661	ValMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAlaSerSerGly	680
Db	1981	GTCACTTCCACGCGGTGCTGCAAAAGAGGATCTGTGGTCTCGCTCGCTCGATCGAGCGG	2040
Qy	681	LeuTyHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla	700
Db	2041	CTTTACCACTCAACCAATGGCGGACGATGCTGTGCAATCAACCGCGGTATCTCCGCG	2100
Qy	701	ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyProAlaValPheValVal	720

Db	2101	GTGACGTTGGGATTTGGTAAGTCTCGCGCGGGTGTGCATACCCAGCGCTCTTTGCTC	2160		Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)
Qy	721	GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpVal	740		Nature 417 (6885), 141-147 (2002)
Db	2161	GGCAGATCGGAGGCGTTACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTA	2220		21996410
Qy	741	LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis	760		12000953
Db	2221	CTGATCAATGATGACGACACCAATACGGAATTTGGGCAACAGCAATCACCGGTGACCCAC	2280		2 (bases 1 to 296500)
Qy	761	AlaAsnLeuAlaArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle	780		Direct Submission
Db	2281	GCGAATTTACGCGGGGTGACATAGCAGCAGCGCGGTGGATGTATACGGGGACATT	2340		Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
Qy	781	GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu	800		On or before Oct 26, 2002 this sequence version replaced
Db	2341	GGTGGTGGCGCGTCCGGATCGCGTCTCGTCCGTGGTGGTCCGTTCGCGGACCTG	2400		gi:20520750, gi:20520810, gi:20520753, gi:20520666, gi:20520757, gi:20520758, gi:20520668, gi:20520759, gi:20520814, gi:20520765, gi:20520767
Qy	801	SerProSerProSerProSerSerProSerProSerProSerProSerProSerProSerPro	820		Location/Qualifiers
Db	2401	AGCCGAGCCGAGCCGAGCAGCTCGCCATCGCGTCCGCTCGCGGCTCGAGTCCA	2460		1. .296500
Qy	821	SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro	840		/organism="Streptomyces coelicolor A3(2)"
Db	2461	TCTCGTCCGCGTCTCCGTCGCGGTCCATCGCCGAGTCCGTCCTGGTCTCGTCAACA	2520		/mol_type="genomic DNA"
Qy	841	SerAlaSerProSerProSerSerProSerProSerProSerProSerProSerProSerProSerPro	860		/strain="A3(2)"
Db	2521	TGGGGTGGCGAGCCGCTCTCGTCACCGAGCCGCTTCGTACCGTTCGTCGCGCG	2580		/db_xref="taxon:100226"
Qy	861	SerProThrProSerSerSerProValSerGlyValIlyValIlyValGlnTyrIlyAsnAen	880		/complement(144. .443)
Db	2581	AGCCCAACCGCTCGTCTCGCGGTGTCGGGTGGGTGAGGTGCAAGTGCAGTATAGATAAT	2640		/genes="SCO6480"
Qy	881	AspSerAlaProGlyAspAsnGlnIlyProGlyLeuGlnValValAsnThrGlySer	900		/notes="synonym: SC9C7.16c"
Db	2641	GATTGGCGCGCGGTGATATCAGATCAAGCCGGGTTTGACAGTGTGTAATACCGGCTCG	2700		complement(144. .443)
Qy	901	SerSerValAspLeuSerThrValThrAlaTyrTyrPheThrArgAspGlyGlySer	920		/genes="SCO6480"
Db	2701	TCTCGGTGGATTTGTCAGCGTGACGTGCGGTACTGTTTACCCGGGATGTTGCTCG	2760		/note="SC9C7.17c, conserved hypothetical protein, len: 348aa; similar to several hypothetical proteins eg. SW.Y08N MYCTU from Mycobacterium tuberculosis (340 aa) fasta scores; opt: 998 z-score: 1321.8, E(): 0, (47.6% identity in 355 aa overlap)."
Qy	921	SerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer	940		/codon_start=1
Db	2761	TCGACACTGGTGTACAACTGTGACTGGCGGCGATCGGGTGGGAATATCCGCGCTCG	2820		/transl_table=11
Qy	941	PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln	956		/product="conserved hypothetical protein"
Db	2821	TTCCGCTCGGTGAACCCGCGAGCGCGGACACCTTACCTTGCAG	2868		/protein_id="CAA22729.1"
LOCUS	SC0393128/c	296500 bp	DNA	linear	/db_xref="GI:4154076"
DEFINITION	Streptomyces coelicolor A3(2) complete genome; segment 25/29.				/db_xref="UniProt/TrEMBL:Q9ZBJ3"
ACCESSION	AL0393128 AL031013 AL031107 AL031350 AL031371 AL031515				/translation="MRWELTSDHARAADTALPGADAVVTRTDTPTDFRGTTFHEVR ARSILNRPVPGASRMPPFTWVYRGCTHACVYCFARKTHYDLDLDTGLGFSQIVVKV ARPEVLRLASRRWQGHVAMGTQYORAEGRYRIMPQGIIEALTRAPFSLITK GTGLRDLDLTLRAARVTEVGSISVGTGDAELWRTVEPGTPAPERLREVVRAALGERG ITCGLMDAPVIFPLGDEPAQLRATVRAAAGTSTVPLVHLRPGAREWMAWIGHRH HPHLVRRYERLYADGAYAPAKWYQRRITRQVHDLAREYIGIGPARAGMPRRVAEPHSRP AEPAPPEPVQLSLI"
VERSION	AL0393128.1	GI:24413907			1800. .2243
KEYWORDS					/genes="SCO6482"
SOURCE	Streptomyces coelicolor A3(2)				/note="synonym: SC9C7.18"
ORGANISM	Streptomyces coelicolor A3(2)				1800. .2243
REFERENCE	1				/genes="SCO6482"
AUTHORS	Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Batenan, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabinowitch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A.				/product="conserved hypothetical protein"

/transl_table=11		DB:	1	Gaps:	18
/product="putative integral membrane protein"		US-09-917-376-1 (1-957) x AP005028 (1-299800)			
/protein_id="BAC69437.1"		QY	9	LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28	
/db_xref="GI:29605371"		DB	175638	CTCACGCTCGAAGAACCCGCATCTTC-----ACGGCCGTGCTCGCGCTG 175682	
/transl_table=11		QY	29	AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46	
/product="putative sugar-binding lipoprotein precursor"		DB	175683	CGCGCGGTCTGCCCGCGGCACCCCGCGGCTGCGCGGAGCGGCCCCACCGCGGACG 175742	
/protein_id="BAC69438.1"		QY	47	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAsp 66	
/db_xref="GI:29605372"		DB	175743	ATCGCGCGGACACGTACAGCTGGAAGAACGCGCGGCTGCGCGGCGGCTTCGTCCCC 175802	
/transl_table=11		QY	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86	
/product="putative LacI-family transcriptional regulator"		DB	175803	GGCATCGTCTTCAACCGCTCCGAGAAGAACCTCGCTTACGCCCGCACCGACATCGCGGC 175862	
/db_xref="GI:29605373"		QY	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyr 106	
/transl_table=11		DB	175863	GCCTACCGCTGGCGCCGAGTCTCTGAAGACCTGAGACGCGCTGCTCGACTCGCTCGCTGG 175922	
/product="putative LacI-family transcriptional regulator"		QY	107	AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126	
/db_xref="GI:29605374"		DB	175923	AGCGACTGGGGGCACACGGGTGTCTGAGGCTCGCTCCGACTCGCTCGACCGCGAACG 175982	
/transl_table=11		QY	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146	
/product="putative LacI-family transcriptional regulator"		DB	175983	GTGTACGCGCGCTGCGCACGTACAGAACAGCTGGGACCGCGGCAACGGTCCGCTGCTC 176042	
/db_xref="GI:29605375"		QY	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166	
/transl_table=11		DB	176043	AGGTCCCGCGACCGGGCGCGAGTCCGACAAAGACCGACCTGCCCTTCAAGCTGCGCGG 176102	
/product="putative LacI-family transcriptional regulator"		QY	167	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186	
/db_xref="GI:29605376"		DB	176103	AACATCGCGCGCGCGGATGGCGGAGCGCTCGCGGTCCGACCGACGAGAACGAGCGGTG 176162	
/transl_table=11		QY	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 206	
/product="putative GntR-family transcriptional regulator"		DB	176163	CTGTACTCGCGCGCGCGCACGGGCTGTGGCGGTGCGCGGCTCGACGACTCGGGGCGCTCC 176222	
/db_xref="GI:29605377"		QY	207	TrpSerGlnMetThrAsnPheProAspValGlyThyTrpIleAlaAsnProThrAspThr 226	
/transl_table=11		DB	176223	TGGTCGCGAGTCAACCGACTTCCCGAACCTCGCGGACCTACGTCGAGGACCGCACACG 176282	
/product="putative LacI-family transcriptional regulator"		QY	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246	
/db_xref="GI:29605378"		DB	176283	AGCGGGTACCGCTCCGACAAACCGGGCATCGTGTGGGTACCTTTCGACGAGTCCGACGGG 176342	
/transl_table=11		QY	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266	
/product="putative proline racemase"		DB	176343	TCGCGGGGAGCTCCACGCGGACGGTGTACTGTCGGGTTCGCGGTCGCGCAAGAGCAACTCCGTC 176402	
/db_xref="GI:29605379"		QY	267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286	
/transl_table=11		DB	176403	TATCGCTCCACGAGCGCGCGCGACCTGTGTGTCGCGCGCGCGCACCGCCACCGGCCAT 176462	
/product="putative LacI-family transcriptional regulator"		QY	287	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306	
/db_xref="GI:29605380"		DB	176463	CTGCCCCACAAAGGGCGGTGCTGGAACGCGCGGCAACGGCTGTCTGTACTCTCGGTACAGCGAC 176522	
/transl_table=11		QY	307	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326	
/product="putative LacI-family transcriptional regulator"		DB	176523	AAGGGCGGACCGTACGACGCGCGGCAAGGACAGCTGTGGCGGTACACGACGAGACCGCGG 176582	
/db_xref="GI:29605381"		QY	327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346	
/transl_table=11		DB	176583	ACCTGGACGAACATCAGCCCGGTTCGCGAGGCGCGACAC-----TACTACGCGCTTC 176633	
/product="putative LacI-family transcriptional regulator"		QY	347	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366	

/transl_table=11		DB:	1	Gaps:	18
/product="putative integral membrane protein"		US-09-917-376-1 (1-957) x AP005028 (1-299800)			
/protein_id="BAC69437.1"		QY	9	LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28	
/db_xref="GI:29605371"		DB	175638	CTCACGCTCGAAGAACCCGCATCTTC-----ACGGCCGTGCTCGCGCTG 175682	
/transl_table=11		QY	29	AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46	
/product="putative sugar-binding lipoprotein precursor"		DB	175683	CGCGCGGTCTGCCCGCGGCACCCCGCGGCTGCGCGGAGCGGCCCCACCGCGGACG 175742	
/protein_id="BAC69438.1"		QY	47	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAsp 66	
/db_xref="GI:29605372"		DB	175743	ATCGCGCGGACACGTACAGCTGGAAGAACGCGCGGCTGCGCGGCGGCTTCGTCCCC 175802	
/transl_table=11		QY	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86	
/product="putative LacI-family transcriptional regulator"		DB	175803	GGCATCGTCTTCAACCGCTCCGAGAAGAACCTCGCTTACGCCCGCACCGACATCGCGGC 175862	
/db_xref="GI:29605373"		QY	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyr 106	
/transl_table=11		DB	175863	GCCTACCGCTGGCGCCGAGTCTCTGAAGACCTGAGACGCGCTGCTCGACTCGCTCGCTGG 175922	
/product="putative LacI-family transcriptional regulator"		QY	107	AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126	
/db_xref="GI:29605374"		DB	175923	AGCGACTGGGGGCACACGGGTGTCTGAGGCTCGCTCCGACTCGCTCGACCGCGAACG 175982	
/transl_table=11		QY	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146	
/product="putative LacI-family transcriptional regulator"		DB	175983	GTGTACGCGCGCTGCGCACGTACAGAACAGCTGGGACCGCGGCAACGGTCCGCTGCTC 176042	
/db_xref="GI:29605375"		QY	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166	
/transl_table=11		DB	176043	AGGTCCCGCGACCGGGCGCGAGTCCGACAAAGACCGACCTGCCCTTCAAGCTGCGCGG 176102	
/product="putative LacI-family transcriptional regulator"		QY	167	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186	
/db_xref="GI:29605376"		DB	176103	AACATCGCGCGCGCGGATGGCGGAGCGCTCGCGGTCCGACCGACGAGAACGAGCGGTG 176162	
/transl_table=11		QY	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 206	
/product="putative GntR-family transcriptional regulator"		DB	176163	CTGTACTCGCGCGCGCGCACGGGCTGTGGCGGTGCGCGGCTCGACGACTCGGGGCGCTCC 176222	
/db_xref="GI:29605377"		QY	207	TrpSerGlnMetThrAsnPheProAspValGlyThyTrpIleAlaAsnProThrAspThr 226	
/transl_table=11		DB	176223	TGGTCGCGAGTCAACCGACTTCCCGAACCTCGCGGACCTACGTCGAGGACCGCACACG 176282	
/product="putative LacI-family transcriptional regulator"		QY	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246	
/db_xref="GI:29605378"		DB	176283	AGCGGGTACCGCTCCGACAAACCGGGCATCGTGTGGGTACCTTTCGACGAGTCCGACGGG 176342	
/transl_table=11		QY	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266	
/product="putative proline racemase"		DB	176343	TCGCGGGGAGCTCCACGCGGACGGTGTACTGTCGGGTTCGCGGTCGCGCAAGAGCAACTCCGTC 176402	
/db_xref="GI:29605379"		QY	267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286	
/transl_table=11		DB	176403	TATCGCTCCACGAGCGCGCGCGACCTGTGTGTCGCGCGCGCGCACCGCCACCGGCCAT 176462	
/product="putative LacI-family transcriptional regulator"		QY	287	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306	
/db_xref="GI:29605380"		DB	176463	CTGCCCCACAAAGGGCGGTGCTGGAACGCGCGGCAACGGCTGTCTGTACTCTCGGTACAGCGAC 176522	
/transl_table=11		QY	307	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326	
/product="putative LacI-family transcriptional regulator"		DB	176523	AAGGGCGGACCGTACGACGCGCGGCAAGGACAGCTGTGGCGGTACACGACGAGACCGCGG 176582	
/db_xref="GI:29605381"		QY	327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346	
/transl_table=11		DB	176583	ACCTGGACGAACATCAGCCCGGTTCGCGAGGCGCGACAC-----TACTACGCGCTTC 176633	
/product="putative LacI-family transcriptional regulator"		QY	347	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366	

/transl_table=11		DB:	1	Gaps:	18
/product="putative integral membrane protein"		US-09-917-376-1 (1-957) x AP005028 (1-299800)			
/protein_id="BAC69437.1"		QY	9	LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28	
/db_xref="GI:29605371"		DB	175638	CTCACGCTCGAAGAACCCGCATCTTC-----ACGGCCGTGCTCGCGCTG 175682	
/transl_table=11		QY	29	AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46	
/product="putative sugar-binding lipoprotein precursor"		DB	175683	CGCGCGGTCTGCCCGCGGCACCCCGCGGCTGCGCGGAGCGGCCCCACCGCGGACG 175742	
/protein_id="BAC69438.1"		QY	47	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAsp 66	
/db_xref="GI:29605372"		DB	175743	ATCGCGCGGACACGTACAGCTGGAAGAACGCGCGGCTGCGCGGCGGCTTCGTCCCC 175802	
/transl_table=11		QY	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86	
/product="putative LacI-family transcriptional regulator"		DB	175803	GGCATCGTCTTCAACCGCTCCGAGAAGAACCTCGCTTACGCCCGCACCGACATCGCGGC 175862	
/db_xref="GI:29605373"		QY	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyr 106	
/transl_table=11		DB	175863	GCCTACCGCTGGCGCCGAGTCTCTGAAGACCTGAGACGCGCTGCTCGACTCGCTCGCTGG 175922	
/product="putative LacI-family transcriptional regulator"		QY	107	AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126	
/db_xref="GI:29605374"		DB	175923	AGCGACTGGGGGCACACGGGTGTCTGAGGCTCGCTCCGACTCGCTCGACCGCGAACG 175982	
/transl_table=11		QY	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146	
/product="putative LacI-family transcriptional regulator"		DB	175983	GTGTACGCGCGCTGCGCACGTACAGAACAGCTGGGACCGCGGCAACGGTCCGCTGCTC 176042	
/db_xref="GI:29605375"		QY	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166	
/transl_table=11		DB	176043	AGGTCCCGCGACCGGGCGCGAGTCCGACAAAGACCGACCTGCCCTTCAAGCTGCGCGG 176102	
/product="putative LacI-family transcriptional regulator"		QY	167	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186	
/db_xref="GI:29605376"		DB	176103	AACATCGCGCGCGCGGATGGCGGAGCGCTCGCGGTCCGACCGACGAGAACGAGCGGTG 176162	
/transl_table=11		QY	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 206	
/product="putative GntR-family transcriptional regulator"		DB	176163	CTGTACTCGCGCGCGCGCACGGGCTGTGGCGGTGCGCGGCTCGACGACTCGGGGCGCTCC 176222	
/db_xref="GI:29605377"		QY	207	TrpSerGlnMetThrAsnPheProAspValGlyThyTrpIleAlaAsnProThrAspThr 226	
/transl_table=11		DB	176223	TGGTCGCGAGTCAACCGACTTCCCGAACCTCGCGGACCTACGTCGAGGACCGCACACG 176282	
/product="putative LacI-family transcriptional regulator"		QY	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246	
/db_xref="GI:29605378"		DB	176283	AGCGGGTACCGCTCCGACAAACCGGGCATCGTGTGGGTACCTTTCGACGAGTCCGACGGG 176342	
/transl_table=11		QY	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266	
/product="putative proline racemase"		DB	176343	TCGCGGGGAGCTCCACGCGGACGGTGTACTGTCGGGTTCGCGGTCGCGCAAGAGCAACTCCGTC 176402	
/db_xref="GI:29605379"		QY	267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286	
/transl_table=11		DB	176403	TATCGCTCCACGAGCGCGCGCGACCTGTGTGTCGCGCGCGCGCACCGCCACCGGCCAT 176462	
/product="putative LacI-family transcriptional regulator"		QY	287	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306	
/db_xref="GI:29605380"		DB	176463	CTGCCCCACAAAGGGCGGTGCTGGAACGCGCGGCAACGGCTGTCTGTACTCTCGGTACAGCGAC 176522	
/transl_table=11		QY	307	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326	
/product="putative LacI-family transcriptional regulator"		DB	176523	AAGGGCGGACCGTACGACGCGCGGCAAGGACAGCTGTGGCGGTACACGACGAGACCGCGG 176582	
/db_xref="GI:29605381"		QY	327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346	
/transl_table=11		DB	176583	ACCTGGACGAACATCAGCCCGGTTCGCGAGGCGCGACAC-----TACTACGCGCTTC 176633	
/product="putative LacI-family transcriptional regulator"		QY	347	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366	

/transl_table=11		DB:	1	Gaps:	18
/product="putative integral membrane protein"		US-09-917-376-1 (1-957) x AP005028 (1-299800)			
/protein_id="BAC69437.1"		QY	9	LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28	
/db_xref="GI:29605371"		DB	175638	CTCACGCTCGAAGAACCCGCATCTTC-----ACGGCCGTGCTCGCGCTG 175682	
/transl_table=11		QY	29	AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46	
/product="putative sugar-binding lipoprotein precursor"		DB	175683	CGCGCGGTCTGCCCGCGGCACCCCGCGGCTGCGCGGAGCGGCCCCACCGCGGACG 175742	
/protein_id="BAC69438.1"		QY	47	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAsp 66	
/db_xref="GI:29605372"		DB	175743	ATCGCGCGGACACGTACAGCTGGAAGAACGCGCGGCTGCGCGGCGGCTTCGTCCCC 175802	
/transl_table=11		QY	67	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86	
/product="putative LacI-family transcriptional regulator"		DB	175803	GGCATCGTCTTCAACCGCTCCGAGAAGAACCTCGCTTACGCCCGCACCGACATCGCGGC 175862	
/db_xref="GI:29605373"		QY	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyr 106	
/transl_table=11		DB	175863	GCCTACCGCTGGCGCCGAGTCTCTGAAGACCTGAGACGCGCTGCTCGACTCGCTCGCTGG 175922	
/product="putative LacI-family transcriptional regulator"		QY	107	AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126	
/db_xref="GI:29605374"		DB	175923	AGCGACTGGGGGCACACGGGTGTCTGAGGCTCGCTCCGACTCGCTCGACCGCGAACG 175982	
/transl_table=11		QY	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146	
/product="putative LacI-family transcriptional regulator"		DB	175983	GTGTACGCGCGCTGCGCACGTACAGAACAGCTGGGACCGCGGCAACGGTCCGCTGCTC 176042	
/db_xref="GI:29605375"		QY	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166	
/transl_table=11		DB	176043	AGGTCCCGCGACCGGGCGCGAGTCCGACAAAGACCGACCTGCCCTTCAAGCTGCGCGG 176102	
/product="putative LacI-family transcriptional regulator"		QY	167	AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186	
/db_xref="GI:29605376"		DB	176103	AACATCGCGCGCGCGGATGGCGGAGCGCTCGCGGTCCGACCGACGAGAACGAGCGGTG 176162	
/transl_table=11		QY	187	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 206	
/product="putative GntR-family transcriptional regulator"		DB	176163	CTGTACTCGCGCGCGCGCACGGGCTGTGGCGGTGCGCGGCTCGACGACTCGGGGCGCTCC 176222	
/db_xref="GI:29605377"		QY	207	TrpSerGlnMetThrAsnPheProAspValGlyThyTrpIleAlaAsnProThrAspThr 226	
/transl_table=11		DB	176223	TGGTCGCGAGTCAACCGACTTCCCGAACCTCGCGGACCTACGTCGAGGACCGCACACG 176282	
/product="putative LacI-family transcriptional regulator"		QY	227	ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246	
/db_xref="GI:29605378"		DB	176283	AGCGGGTACCGCTCCGACAAACCGGGCATCGTGTGGGTACCTTTCGACGAGTCCGACGGG 176342	
/transl_table=11		QY	247	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266	
/product="putative proline racemase"		DB	176343	TCGCGGGGAGCTCCACGCGGACGGTGTACTGTCGGGTTCGCGGTCGCGCAAGAGCAACTCCGTC 176402	
/db_xref="GI:29605379"		QY	267	PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286	
/transl_table=11		DB	176403	TATCGCTCCACGAGCGCGCGCGACCTGTGTGTCGCGCGCGCGCACCGCCACCGGCCAT 176462	
/product="putative LacI-family transcriptional regulator"		QY	287	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306	
/db_xref="GI:29605380"		DB	176463	CTGCCCCACAAAGGGCGGTGCTGGAACGCGCGGCAACGGCTGTCTGTACTCTCGGTACAGCGAC 176522	
/transl_table=11		QY	307	ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326	
/product="putative LacI-family transcriptional regulator"		DB	176523	AAGGGCGGACCGTACGACGCGCGGCAAGGACAGCTGTGGCGGTACACGACGAGACCGCGG 176582	
/db_xref="GI:29605381"		QY	327	ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346	
/transl_table=11		DB	176583	ACCTGGACGAACATCAGCCCGGTTCGCGAGGCGCGACAC-----TACTACGCGCTTC 176633	
/product="putative LacI-family transcriptional regulator"		QY	347	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366	

/transl_table=11		DB:	1	Gaps:	18

Alignment Scores:
Pred. No.: 5, 27e-65
Score: 2487.00
Percent Similarity: 64.77%
Best Local Similarity: 50.78%
Query Match: 48.43%

FNGAYTSKNTPTPTSFSLNGVTCALAD*			
sig_peptide	1..69		
mat_peptide	70..2820		
/product="unnamed"			
ORIGIN			
Alignment Scores:			
Pred. No.:	1.74e-57	Length:	2823
Score:	2174.00	Matches:	430
Percent Similarity:	59.48%	Conservative:	141
Best Local Similarity:	44.79%	Mismatches:	311
Query Match:	42.34%	Indels:	78
DB:	6	Gaps:	16
US-09-917-376-1 (1-957) x AX565635 (1-2823)			
QY	10	ThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAla	29
DB	13	ACCGTGGTGGCCACAGCACTAACTGCGCTGCTCGCCAGCAGCTGCGGTGGCATCGCGCC	72
QY	30	AlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThr	49
DB	73	ACATCA-----ACCCCATGCGCAGTAACCTGCACCGCCCTCTTCTCATCATCACT	123
QY	50	GlnProTyrThrTTPSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleVal	69
DB	124	TCCGGCTATAGCTGTGCGAACCTTGAGATTGTGGCGGCGGTTTGTGCTGGAATTGTC	183
QY	70	PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg	89
DB	184	TTTAAACCAAGAACCCAGCGTGTGGTGTATGTCGCGCACAGATATTGGTGAGCGGTACCGG	243
QY	90	TTPAspAlaAlaAsnGlyArgTTPProLeuLeuAspTTPValGlyTTPAsnAsnTTP	109
DB	244	CTCATGACTCCACCGGGGTGGATCTCTTACGGACCATATTGGTTGGGATGACTGG	303
QY	110	GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTTPAla	129
DB	304	AGCCACTCGGAATTCTGCTTGGCCACTGACCTGTTGACACCAACCGCGGTGACTTT	363
QY	130	AlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyValAlaIleLeuArgSer	149
DB	364	GCAGCAGCACGTACTCCATGGAAGTGGGATCCCAAAACGGTGTATTTGGCTTCGCG	423
QY	150	AspGlnGlyAlaThrTTPGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetPro	169
DB	424	GACAAAGCGGAGACTGGGAAAGACGATGCTGCGGTCTCGTGTGGAGGCAACATGCCG	483
QY	170	GlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPhe	189
DB	484	GGTGGCGGGATGGGAGAACGCTCGCGGTGACCCCAACAATAAAGTCTCTGTACTTT	543
QY	190	GlyAlaProSerGlyLysGlyLeuTTPArgSerThrAspSerGlyAlaThrTTPSerGln	209
DB	544	GGTGGGAATCGGGTAATGGACTATGGAAGTCAACTGACTACGGGAAACCTGGGGCAA	603
QY	210	MetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly---	228
DB	604	GTGACGAGCTTCCCAATGCTGTAACTACGTGCC-----GATGCTTCTGTGTGG	654
QY	229	TyrGlnSerAspIleGlnGlyValValTTPValAlaPheAspLysSerSerSerLeu	248
DB	655	TACACCGGCAAAACCAAGGTGTGTGGGTAACTTTTGACCCCATCATCAGCGAAAGCT	714
QY	249	GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTTP	268
DB	715	GGTCAGACAAACACAAACGATCTAGCTAGGTGTAGCAGCAAGCAGAACACGTTTATCGA	774
QY	269	SerArgAspGlyAlaThrTTPGlnAlaValProGlyValAlaProThrGlyPheIlePro	288
DB	775	TCTACCGATGGCGGGGCACATGGACGCGTCTCTGGACCAACCCCGGATTCTCTGGCG	834
QY	289	HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly	308

Db	1900	GCCTCAGCTTTCACGGGGTTTCCACCACTGGA-----AACGTTGCTTTCCGTGCGGTG	1953
Qy	667	ProGlyIysGluGlyAspLeuThrLeuAla-----AlaSerSerGlyLeu	681
Db	1954	CCGGCGCGTCAAGGTCTACTCTGCTTGGTGGAGTATCGGAAGCACCTATGGGATG	2013
Qy	682	TyrHisSerThrAsnGlySerSerThrSerAlaIleThrGlyValSerSerAlaVal	701
Db	2014	TGGCGGTCCTCATGATGGCGGAAGACTGGACCAAGTATCTGGAGTTCAGGAAGCGAT	2073
Qy	702	AsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGly	721
Db	2074	GCCGTTGGTTTCGTTAAGCAACCAAGTAGTAGCGGATACCCAGTGATTTTTCACATCCGCC	2133
Qy	722	ThrIleGlyGlyValThrGlyAlaThrArgSerAspAspCysGlyThrThrTrpValLeu	741
Db	2134	AAGATTGATGAGTTCGAGGAATTTTCGGTCTGACGATGAAGGCAAGACCTGGAACGC	2193
Qy	742	IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla	761
Db	2194	ATCAACGACACCAACCAACGATGATTTGGACTGGCGCATCAATCACCGGTGACCCAGAT	2253
Qy	762	AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGly	781
Db	2254	GTCGAAGGCGCGTCTACATTGGTACCAACGTCGCGGCAATTATCGTTGGTGAT-----	2307
Qy	782	GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer	801
Db	2307	-----	2307
Qy	802	ProSerProSerProSerSerProSerProSerProSerProSerSerProSer	821
Db	2308	-----TCTCCACTCCCCCGCCCCCGCTCAAGTAGCGGACGCGCATCG	2352
Qy	822	SerSerProSerProSerProSerProSerProSerProSerArgSerProSerPro	840
Db	2353	GGAATACAGGGGCCACACGCGCTGAACCCCTGCACCGATCAACACTGGCGGG	2412
Qy	841	SerAlaSerProSerProSerProSerProSerProSerProSerProSerSerPro	860
Db	2413	TCCACTGATACGGGCGAACTTCAGAGCGGGAACAGGAGTGGATCTCGAAGCCGCACT	2472
Qy	861	SerProThrProSerSerSerProValSerGlyGlyValValValGlnTyrLysAsnAsn	880
Db	2473	GACTCGGGGAACACTCTCTCGGGAATACCAAGTTCATGCAAGGTTCCGTATTCACGACA	2532
Qy	881	AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAlaSerThrGlySer	900
Db	2533	GACTGGGCGCGGC-----TTCACTGGGCGGCGACCATCACTAAACAGTCCCCC	2583
Qy	901	SerSerValAspLeuSerThrValThrValArgTyr-----	912
Db	2584	AGCCCAATTAACGGGTGACGCTGGCGTCTACTTACCCTTCGGGCGACCATCAGTTCC	2643
Qy	913	---TrpPheThrArgAsp-----GlyGlySerSerThrLeuValTyrAsnCysAspTrp	929
Db	2644	GTGTGGTCAGCAACTCAGACTCTGTCTGTGGTCGCAATGTTGTGCTCAAGAACTCTGTTGG	2703
Qy	930	AlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro	949
Db	2704	-----AATCCACAGTACCT	2718
RESULT	6		
LOCUS	AE007608	12732 bp	DNA linear BCT 01-JUN-2004
DEFINITION	Clostridium acetobutylicum ATCC 824 section 96 of 356 of the complete genome.		
ACCESSION	AE007608	AE001437	
VERSION	AE007608.1	GI:15023819	
KEYWORDS			
SOURCE	Clostridium acetobutylicum ATCC 824		
ORGANISM	Clostridium acetobutylicum ATCC 824		
	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		

REFERENCE	Clostridium.		
AUTHORS	1 (bases 1 to 12732) Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q., Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I., Tatusov,R.L., Sabathe,F., Koonin,E.V., Doolittle,W., Soucaille,P., Daly,M.J., Bennett,G.N., Dounin,E.V. and Smith,D.R. Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum J. Bacteriol. 183 (16), 4823-4838 (2001) 21359325		
TITLE	1. 12732		
JOURNAL	/organism="Clostridium acetobutylicum ATCC 824"		
MEDLINE	/mol_type="genomic DNA"		
PUBMED	/strain="ATCC 824"		
REFERENCE	2 (bases 1 to 12732) Childress,D., Zeng,Q. and Smith,D.R. Direct Submission Submitted (24-JUL-2001) GTC Sequencing Center Production, Finishing, and Bioinformatics Teams, Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02453-8443, USA Location/Qualifiers		
AUTHORS	1. 12732		
TITLE	/organism="Clostridium acetobutylicum ATCC 824"		
JOURNAL	/mol_type="genomic DNA"		
MEDLINE	/strain="ATCC 824"		
PUBMED	/db_xref="ATCC:824"		
REFERENCE	73. .2592		
AUTHORS	/gene="CAC0919"		
TITLE	73. .2592		
JOURNAL	/gene="CAC0919"		
MEDLINE	/gene="CAC0919"		
PUBMED	/codon_start=1		
REFERENCE	/transl_table=11		
AUTHORS	/product="Probably secreted sialidase; several ASP-boxes and dockerin domain"		
TITLE	/protein_id="AAK78895.1"		
JOURNAL	/db_xref="GI:15023820"		
MEDLINE	/translation="MNKRIVSMVAGLSIIFTTGVTHISAAANKAAASVOSQYKWDN AKICAGGVPAVIFNKTKBLIYARTDMGAYRWDKANNKWIPIITDGFSDMTLGCES IATDPIDNVRVIAAGLYTNDWDENAVILSDQKGTWKRYQLPFKVGNNPGRNMG ERLQIDPNDKILYLGBSGNGLKESDYGGTWSKVDNFPDGTGVQDPQNEYTADKV GVWETFPDSTGTGSGPTQMTVGAADTKGNIYTNDDGKTWSAVKGGPKGYLPHHG ILASDGMILYISYNTSGPYDGSDDGVKINTKTGEMTNITPPAVGDTKSGFGGISVDA QKNPVVATLNRWMPDEEIRYSTDAGTKWPIWDMNGYPNRTLTGTLNLDYSAPWLDW KGTVPDPLVPLKLGMMGLLEIDPFNSDRMPYGTGATLYGTDLLTNMDKKNVDISV KANGIECAVNDVVPTKGAQLLSAVGDDCGFYHDDITKVPKMMTTPNFSATTSIDY AESVPNFVRVGNVDTSKNQDKDCGISYDGKKNWFSAGSNISGVYKACTVAAGADAK TIWSPREGAANAATDNGNKTWPCSGIPQGAUKASDRVNPKFVGLNGKPYISTDA DATTQSSQTGLPTKGIFKTVIGHGDIWAGKDGMLHSTDSGATFTKSGVSDAS DTVGLGSKTDDGYPAIYMDATIDGTAGIFRSDDEGATWIRINDAHOYGSPPDYCTG DPNKYGVFVGTNGRGIYVGDIDGSOPTTPSPVTLGDVNDGDIINGRDVWVLTQYIAG KDVTIDKKAADVNGDGTVNGRDLMTLRQYISGKITSFSSSTK"		
gene	complement(2634. .2978)		
CDS	/gene="CAC0920"		
	complement(2634. .2978)		
	/gene="CAC0920"		
	/codon_start=1		
	/transl_table=11		
	/product="Protein related to MIFH/DOPD protein family, function in bacteria is unknown"		
	/protein_id="AAK78896.1"		
	/db_xref="GI:15023821"		
	/translation="MPFINSTVTVKLDKEKSTLVELKGIIELLPKGSEDWLMVGFK DNYPLYPKGEKKDAAFIEVKIFGSAKSNKLYSAISSLMENELSIIPKDCIYITVD EVSNWAWNGSLF"		
gene	3104. .4252		
CDS	/gene="CAC0921"		
	3104. .4252		
	/gene="CAC0921"		
	/codon_start=1		
	/transl_table=11		
	/product="Metal-dependent amidohydrolase"		
	/protein_id="AAK78897.1"		
	/db_xref="GI:15023822"		
	/translation="MITIEKOEPLKLLIGYRRELHKHPHLSMKEVETTKRIKRWLYE NDIKVLDLPFIEVGVAEIEGEHEGKTVAIRADILDALPIEKELTPELSSVNNKGINHAGG		

gene	HDEHTAAILGTAIILSKRAEIIYGRVRIIFOPGEETGKGSQVYIEBGLVRDVCIFGM	
	HNKPLPVTGIIIGSGLMWSVDREIKVIGSGGHAGIPNKCIDPIVASOIVASQMT	
CDS	VYRSISIESVVISITRFSNVTWVSDSAREGTYRTFNNIRISIRNLMERTAK	
	GIARALGKIFKMFSPVLENDVRFKDLCEIAAEDGRTYKVEARNIGGEDFSFYQ	
gene	TVPSFFVMGVGDTGEWHKANYNLKEDALVAANYFSDLALVKLSKK"	
	4594. .4863	
CDS	/gene="CAC0922"	
	4594. .4863	
gene	/product="Uncharacterized small conserved protein, homolog of yfjA/yukE B subtilis"	
	/db_xref="GI:15023823"	
CDS	/protein_id="AAK78898.1"	
	5056. .6348	
gene	/gene="CAC0923"	
	5056. .6348	
CDS	/gene="CAC0923"	
	/codon_start=1	
gene	/transl_table=11	
	/product="Predicted membrane protein; CF_1 family"	
CDS	/protein_id="AAK78899.1"	
	/db_xref="GI:15023824"	
gene	/translation="MQSLKVMRNGLLKINEAKQGLKSAVDKFSLNKRSRSTIEIKN	
	LKRELVKQLEIGETKINSINIDYVGRFTDVKCKASKLTSGYTKYRKGLNLFGE	
CDS	EVERYVGGIGVPIIDAVESVDRNKFVLTQTTTLETGAIGVCIVTLPEAGITWKA	
	WDALSAAGVSDNIGTAASAIYORKVERKEEDSSGYDLFQSIFFGAGWYSAKILG	
gene	VKNPSYWCSSGKTLVDRTSATVTLNVGHSVTSLGKDVEKLPALKKAESSANKTLEA	
	ANNAEKIHMDAMATKINGINNAKKEVVDLKKLKNVGTNPFEPKQFYDKNLL	
CDS	KDKGNKLQLOQLLIQDQLETFEKTNTLNANKYKNSIRDNLKNRYDRIKASLNIPA	
	GGTLNTEDVQGSFHFYPPGGEIKPIFEIKVVLGGR"	
gene	6345. .6923	
	/gene="CAC0924"	
CDS	6345. .6923	
	/gene="CAC0924"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="Predicted membrane protein"	
	/protein_id="AAK78900.1"	
gene	/db_xref="GI:15023825"	
	/translation="MIRFOKRAIFPLTISILIIICGDIKIISFGIEFDIILCNLFFI	
CDS	GLGIIGFRIYRWYKQIKYKYTELLKEDRLDKIEYIYDIERVGVVYNTAFYIDGVKC	
	KLRFPOYCKNDNVEYULEPIYKNGEFPKDFIADVADKNDYKKSISTESTNEITCEBEK	
gene	AEKKRRKRLNYSVNSRMLAKYBEEAKNK"	
	7003. .8043	
CDS	/gene="CAC0925"	
	7003. .8043	
gene	/gene="CAC0925"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="TPR-repeat-containing protein"	
gene	/protein_id="AAK78901.1"	
	/db_xref="GI:15023826"	
CDS	/translation="MENENVLRAQLNRLNGLIFIERNAGDEALKYFNKAIKEDENFKE	
	AVLNKAALCLANNKYDDAVSCYDKLIIKYVNEGKAYFGNGVWFFYKDKNIKKAILEYN	
gene	RSIYLGEMSEGVYNALCMNESIGELDDAIKWLNRALIINDONIITLNKKAVALIKIK	
	KFYALEECNRVLKIEVDNEEVTHFKAMLLGEMGKEIKAFETISRGREALLGEMTLCY	
CDS	DKVALFEKQKEFKALCEKTEKALLFDGNNVLTLLKGGDLTLTKLIDKIDKASTYDKILE	
	FHNMMGVFPAKANLCMLLGEHEDAEFFKEIIERLSGSESVLINSYYYRNLNLKKWA	
gene	KTRKQRLIRKQ"	
	7998. .8357	
CDS	/gene="CAC0926"	
	7998. .8357	
gene	/gene="CAC0926"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="TPR-repeat-containing protein"	
gene	/protein_id="AAK78902.1"	
	/db_xref="GI:15023827"	
CDS	/translation="MGKNBEAKYKEATKKYNFLIIKYPYDVKLNFFKANCRLDIED	
	YKABELYEIYIDLNDALLEVYLMRARNRISLNKYBAAKEDLNTVTKLNPAYRKVIEL	
gene	8379. .8675	
	/gene="CAC0927"	
CDS	8379. .8675	
	/gene="CAC0927"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="Hypothetical protein, CF-2 family"	
	/protein_id="AAK78903.1"	
gene	/db_xref="GI:15023828"	
	/translation="MNSDEIRIDEKADKDVIKDLKAVGEKVKETYNLIHLSNRHIEDG	
CDS	IKGSAVDSITTEYNKIYKVMVKEMQDNMGNDVAAVNNIVSSFDAEDKMIKTKFYR"	
	8713. .9012	
gene	/gene="CAC0928"	
	8713. .9012	
CDS	/gene="CAC0928"	
	/codon_start=1	
gene	/transl_table=11	
	/product="Uncharacterized small conserved protein, homolog of yfjA/yukE B subtilis"	
CDS	/protein_id="AAK78904.1"	
	/db_xref="GI:15023829"	
gene	/translation="MVDRCDIEIDASIFEETIKYVSDSKVRLHDLCDLENELRMED	
	TWEGDAKEFDSTFPGFYSAMKKDCDMLBELIKELKLVKTSFESLDTMKNLDKK"	
CDS	9712. .10335	
	/gene="CAC0929"	
gene	9712. .10335	
	/gene="CAC0929"	
CDS	/codon_start=1	
	/transl_table=11	
gene	/product="SAM-dependent methyltransferase"	
	/protein_id="AAK78905.1"	
CDS	/db_xref="GI:15023830"	
	/translation="MGSMQYFDNIAASWNVIRSEYFEERLYKVLKSVIKDKITADF	
gene	GCGTGFISALASDANNFGLDNNMLRKLSAHDKLNINYLKSLTSVTLFDE	
	SIDAVFINNALHHI KDAEKAIKEMRYVLKDGTVVISDVSEHNGKWARENFDEWLGF	
CDS	SNEQIDMWLKVGVFKNIDIENTDLCKGTSRSGHTETSIFIACAAK"	
	10353. .11498	
gene	/gene="CAC0930"	
	/notes="metB"	
CDS	10353. .11498	
	/gene="CAC0930"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="TPR-repeat-containing protein"	
	/protein_id="AAK78901.1"	
gene	/db_xref="GI:15023825"	
	/translation="MIRFOKRAIFPLTISILIIICGDIKIISFGIEFDIILCNLFFI	
CDS	GLGIIGFRIYRWYKQIKYKYTELLKEDRLDKIEYIYDIERVGVVYNTAFYIDGVKC	
	KLRFPOYCKNDNVEYULEPIYKNGEFPKDFIADVADKNDYKKSISTESTNEITCEBEK	
gene	AEKKRRKRLNYSVNSRMLAKYBEEAKNK"	
	7003. .8043	
CDS	/gene="CAC0925"	
	7003. .8043	
gene	/gene="CAC0925"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="TPR-repeat-containing protein"	
gene	/protein_id="AAK78901.1"	
	/db_xref="GI:15023826"	
CDS	/translation="MENENVLRAQLNRLNGLIFIERNAGDEALKYFNKAIKEDENFKE	
	AVLNKAALCLANNKYDDAVSCYDKLIIKYVNEGKAYFGNGVWFFYKDKNIKKAILEYN	
gene	RSIYLGEMSEGVYNALCMNESIGELDDAIKWLNRALIINDONIITLNKKAVALIKIK	
	KFYALEECNRVLKIEVDNEEVTHFKAMLLGEMGKEIKAFETISRGREALLGEMTLCY	
CDS	DKVALFEKQKEFKALCEKTEKALLFDGNNVLTLLKGGDLTLTKLIDKIDKASTYDKILE	
	FHNMMGVFPAKANLCMLLGEHEDAEFFKEIIERLSGSESVLINSYYYRNLNLKKWA	
gene	KTRKQRLIRKQ"	
	7998. .8357	
CDS	/gene="CAC0926"	
	7998. .8357	
gene	/gene="CAC0926"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="TPR-repeat-containing protein"	
gene	/protein_id="AAK78902.1"	
	/db_xref="GI:15023827"	
CDS	/translation="MGKNBEAKYKEATKKYNFLIIKYPYDVKLNFFKANCRLDIED	
	YKABELYEIYIDLNDALLEVYLMRARNRISLNKYBAAKEDLNTVTKLNPAYRKVIEL	
gene	8379. .8675	
	/gene="CAC0927"	
CDS	8379. .8675	
	/gene="CAC0927"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="Hypothetical protein, CF-2 family"	
	/protein_id="AAK78903.1"	
gene	/db_xref="GI:15023828"	
	/translation="MNSDEIRIDEKADKDVIKDLKAVGEKVKETYNLIHLSNRHIEDG	
CDS	IKGSAVDSITTEYNKIYKVMVKEMQDNMGNDVAAVNNIVSSFDAEDKMIKTKFYR"	
	8713. .9012	
gene	/gene="CAC0928"	
	8713. .9012	
CDS	/gene="CAC0928"	
	/codon_start=1	
gene	/transl_table=11	
	/product="Uncharacterized small conserved protein, homolog of yfjA/yukE B subtilis"	
CDS	/protein_id="AAK78904.1"	
	/db_xref="GI:15023829"	
gene	/translation="MVDRCDIEIDASIFEETIKYVSDSKVRLHDLCDLENELRMED	
	TWEGDAKEFDSTFPGFYSAMKKDCDMLBELIKELKLVKTSFESLDTMKNLDKK"	
CDS	9712. .10335	
	/gene="CAC0929"	
gene	9712. .10335	
	/gene="CAC0929"	
CDS	/codon_start=1	
	/transl_table=11	
gene	/product="SAM-dependent methyltransferase"	
	/protein_id="AAK78905.1"	
CDS	/db_xref="GI:15023830"	
	/translation="MGSMQYFDNIAASWNVIRSEYFEERLYKVLKSVIKDKITADF	
gene	GCGTGFISALASDANNFGLDNNMLRKLSAHDKLNINYLKSLTSVTLFDE	
	SIDAVFINNALHHI KDAEKAIKEMRYVLKDGTVVISDVSEHNGKWARENFDEWLGF	
CDS	SNEQIDMWLKVGVFKNIDIENTDLCKGTSRSGHTETSIFIACAAK"	
	10353. .11498	
gene	/gene="CAC0930"	
	/notes="metB"	
CDS	10353. .11498	
	/gene="CAC0930"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="TPR-repeat-containing protein"	
	/protein_id="AAK78901.1"	
gene	/db_xref="GI:15023825"	
	/translation="MIRFOKRAIFPLTISILIIICGDIKIISFGIEFDIILCNLFFI	
CDS	GLGIIGFRIYRWYKQIKYKYTELLKEDRLDKIEYIYDIERVGVVYNTAFYIDGVKC	
	KLRFPOYCKNDNVEYULEPIYKNGEFPKDFIADVADKNDYKKSISTESTNEITCEBEK	
gene	AEKKRRKRLNYSVNSRMLAKYBEEAKNK"	
	7003. .8043	
CDS	/gene="CAC0925"	
	7003. .8043	
gene	/gene="CAC0925"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="TPR-repeat-containing protein"	
gene	/protein_id="AAK78901.1"	
	/db_xref="GI:15023826"	
CDS	/translation="MENENVLRAQLNRLNGLIFIERNAGDEALKYFNKAIKEDENFKE	
	AVLNKAALCLANNKYDDAVSCYDKLIIKYVNEGKAYFGNGVWFFYKDKNIKKAILEYN	
gene	RSIYLGEMSEGVYNALCMNESIGELDDAIKWLNRALIINDONIITLNKKAVALIKIK	
	KFYALEECNRVLKIEVDNEEVTHFKAMLLGEMGKEIKAFETISRGREALLGEMTLCY	
CDS	DKVALFEKQKEFKALCEKTEKALLFDGNNVLTLLKGGDLTLTKLIDKIDKASTYDKILE	
	FHNMMGVFPAKANLCMLLGEHEDAEFFKEIIERLSGSESVLINSYYYRNLNLKKWA	
gene	KTRKQRLIRKQ"	
	7998. .8357	
CDS	/gene="CAC0926"	
	7998. .8357	
gene	/gene="CAC0926"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="TPR-repeat-containing protein"	
gene	/protein_id="AAK78902.1"	
	/db_xref="GI:15023827"	
CDS	/translation="MGKNBEAKYKEATKKYNFLIIKYPYDVKLNFFKANCRLDIED	
	YKABELYEIYIDLNDALLEVYLMRARNRISLNKYBAAKEDLNTVTKLNPAYRKVIEL	
gene	8379. .8675	
	/gene="CAC0927"	
CDS	8379. .8675	
	/gene="CAC0927"	
gene	/codon_start=1	

Db 295 GCATATAGATGGCAACAGCTAATAATAAGTGGATTCCAAATTACTGAT-----GGATTT 348
QY 107 AenAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAaspProIleAsnThrAsnLys 126
Db 349 TCTGATTGGCAAAATGCTTGGCTGCGAAAGTATAGCAACTGACCCCAATAGACATAATCGT 408
QY 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146
Db 409 GTTTATATAGCAGCAGGTTGTACAAATGATTGGCAAGATGAAATGCATATATCTT 468
QY 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166
Db 469 AGTTCACAGGATAAAGGTAAACACATGGAAGATACCAACTCCATTTTAAAGGTAGAGGT 528
QY 167 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186
Db 529 AATATGCCGGGAAGAAATATGGGAGAAAGACTTCAATAGATCCCTAATAGATATAAATA 588
QY 187 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206
Db 589 CTTTATCTTGAGCTAGAGTGGTAAATGGACTTTGGAAAGTGAAGACTATGGACAACA 648
QY 207 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226
Db 649 TGGTCTAAGTTGATAATTTCCCTCATACAGCGACTATGTACAAGATCCTCAAAATGAA 708
QY 227 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 246
Db 709 -----TATACAGCAGATAAAGTTGGAGTAGTTTGGGAACCTTTTGATCCTTCAACAGGA 762
QY 247 SerLeuGlyGlnAlaSerIleThrIlePheValGlyValAlaAasp---ProAsnAsnPro 265
Db 763 ACAAAAGGAAGCCCAACTCAAACTATGATGTGTGGTGACAGAGATAAGACAGAAACAAT 822
QY 266 ValPheTrpSerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 285
Db 823 ATTACGTTACTAATGATGGAGGAAGACATGAGTGCGAGTTAAAGGACAGCCAAAGGA 882
QY 286 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 305
Db 883 TATCTTCCACATCATGTTATCTT---GCTTCCGATGGTATGCTTTATATATCATATAGC 939
QY 306 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 325
Db 940 AATACTTCGGACCATATAGCGGCTCAGATGGACAAAGTTTGAAGATATAAATACCAAAACA 999
QY 326 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 345
Db 1000 GGAGAAATGGCAAAACATACTCACTGCAGTAGGTGACACTAAATCT-----GGT 1050
QY 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365
Db 1051 TTTGGAGGAATATCAGTAGATGCACAAATCCAAATAAATGTTGTAGTTGCAACTTTAAAT 1110
QY 366 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 385
Db 1111 AGATGGTGGCCGGATGAAGAAATATACAGAAGTACAGATGCAGGTAAAGACCTGGAAACCT 1170
QY 386 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 405
Db 1171 ATATGGGATGGAATGGATATCCAAACCGTACTTTGGGATATACTTAGATTACTCAGCA 1230
QY 406 GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSerProLysLeu 424
Db 1231 CAACCTTGGCTTGGTTGGGTGAAGACCGGGGTAAACACACACAGATCCATTTAGTTAAATTA 1290
QY 425 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 444
Db 1291 GGCTGGATGATGGGAGACTTAGAAAATTGATCCATTTAACTCGGATAGAATGTTCTACGGT 1350
QY 445 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIle 464
|||||

Db 1351 ACAGGTGCAACTCTTTATGGAACTGATGATTTTGACTAAATTTGGGATAAAGTAAAGATGTA 1410
QY 465 HisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro 484
Db 1411 GACATCTCTGTAAAGCAAAATGGAATTTGAAGAATGTGAGTAATAATGATGTGTAGTACCA 1470
QY 485 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAasp 504
Db 1471 ACAAAAGGCGCAACACTTTTAAAGCGCTGTTGGAGATGACTCGCGGATTTTATCATGATGAC 1530
QY 505 ValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524
Db 1531 ATAACCTAAAGTTTCCAAGTAAATGATGACAACAACCTTAACCTTTAGTGTCTGCTCAACAAGCAT 1590
QY 525 AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSer 544
Db 1591 GATTATGCGAAAGTGTCTTAACCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1650
QY 545 GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln--- 563
Db 1651 AATCAACAGATAAAGATTTGGAATTTCTACGATGGTGGTAAAGAACTGGTTCTCGACT 1710
QY 564 GlySerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySer 583
Db 1711 GGAAGCAATATATCAGGAGTTTTATAAAGCAGGAACCTGTAGCAGCTGGTGTGATGCTAAA 1770
QY 584 ArgPheValTrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGly 603
Db 1771 ACTATAGTATGTTCTCCA---GAGGAAGGTGCTTAATGCAGCATATATGACAGATAATAGT 1827
QY 604 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 623
Db 1828 AATAAATGGACACCGCTGTTCTGGTCTTCCACAGGAGCAAGGTAGATCAGACCGGTGA 1887
QY 624 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 643
Db 1888 AATCAAAGAAATTTCTATGTTCTTAAATGGTAAATTTCTATATAGTATGATCTGATGAGGT 1947
QY 644 ValThrPhe---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet 662
Db 1948 GCAACATTTTACTCAAAGTTTCAACAACAGGTCTTCTTACAAAGGCAAA---GGAATC--- 2001
QY 663 PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAla---AlaSerSerGlyLeu 681
Db 2002 TTTAGACTGTTATAGTTCATGAGTGTATTTGGATTTGCTGGAGGAAAGACGGATTA 2061
QY 682 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaVal 701
Db 2062 TGGCATTCAACTGATTCTCGTGCAACATTCACAAAAGTATCTGTTGTTGATGCTTCGTAT 2121
QY 702 AsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGly 721
Db 2122 ACAGTAGGTCTTGGAAATCCAAAACAGATGATGGATATCCAGCTATATACATGGATGCT 2181
QY 722 ThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeu 741
Db 2182 ACAATTTGAGAACACCGCGGAATATTCGGTTTCAGATGATGAAGTGCACCTTGGGTAGA 2241
QY 742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761
Db 2242 ATAAATGATGATGCTCACCAATATGTTTCCCGACACTTACTGTCATACTGGAGATCCTAAC 2301
QY 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGly 781
Db 2302 AAATATGACGAGTTTTGTGGAACTTAATGGAAGAGGTATTGTTTATGATGAGACATAGAT 2361
QY 782 GlyAlaProSerGlySerProSerProSerValSer 793
Db 2362 GGTTCCT---CAACCTACACCTACACCAAGTGTACT 2394
RESULT 7
AJ585344
LOCUS

AJ585344 2950 bp DNA linear BCT 11-OCT-2003

QY 404 SerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLys 423
Db 1520 TCCGCAGCACCGTGGCTGGATTGGGAACTGAGAAACAGCTGCCGCGAAATCAATCCGAAA 1579
QY 424 LeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAenSerAspArgMetLeuTyr 443
Db 1580 CTGGATGGATGATAGGTGACATAGAGATTGACCGGTTTAAATTCGACCGCATGTAT 1639
QY 444 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGln 463
Db 1640 GTTACCGGTGCACACTATGTTGTGACAACTTACTGACTGGACAGAGCGCGCAA 1699
QY 464 IleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSer 483
Db 1700 GTAAAAATCGAGTTAAAGCTACCGGAATAGAGAAATGTCGGTATTAGACCTGTTAAGC 1759
QY 484 ProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla 503
Db 1760 CGCCCGAGGGTGCACCGCTTGAAGTGCAGTTGGCGACCTTGTCTGCTTTCATGAT 1819
QY 504 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 523
Db 1820 GACCTGAAAGTTGGTCCGAAAAAATGACGTT---CCTTCTTATTTCTTCAGGTACGGA 1876
QY 524 ValAspTyrAlaGluLeuAsnPro-----SerIleIleValArgAlaGlySerPhe 540
Db 1877 ATTGATTATCGGACCTTGTCCGAACTTATGGCAATGTTGGTTCGAAAGCTGATTTGTAT 1936
QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560
Db 1937 GATGTA-----AAGAAGATTCTTCTCTATATCGGAGGAAGGAAT 1978
QY 561 TrpPheGln--GlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSer 579
Db 1979 TGGTTCACGCCACCTAATAGACACCAACCTCGTAGCGCGGGTTCGGTTCGCCGTGCA 2038
QY 580 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 599
Db 2039 GCCGATGCMAATCAGTTATTGGACCG---GAAATGCAAGTCTCGCAGTT---ACA 2092
QY 600 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 619
Db 2093 ACGGACCAACGGAACTCATGGAAGTGTGTACAAATCTTGGTATGGTGGTGGTGGCA 2152
QY 620 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 639
Db 2153 TCCGACCGTGTGAACGGTAAAAAATTTACGCAATTTCTATAACCGCAAAATTTCTATAAAGC 2212
QY 640 ThrAspGlyGlyValThrPheGlnProVal---AlaAlaGlyLeuProSerSerGlyAla 658
Db 2213 ACGGACGGTGGATTAACTTTTACCATACAAGGCACCGCAGCTTCCCAAGTCCGTTAAC 2272
QY 659 ValGlyValMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAla--- 677
Db 2273 -----AAGATAAAACCGTACCGGCAAGGAAGGACATGATGCTGCTGCAAGA 2323
QY 678 SerSerGlyLeuTyrHisSerThrAsnGlyLysSerSerTrpSerAlaIleThrGlyVal 697
Db 2324 GAAGCGGATTTGGAGGTCCCACTGACGGTGGATATACGTTTGAGAAACTCTCCAATGTT 2383
QY 698 SerSerAlaValAsnValGlyPheGlyLysSerAlaProGlyLysSerSerTrpProAlaVal 717
Db 2384 GACACGCTCATGTGTAGCTTCGGAAGCGCAGCACCGGACAGGATTACATGCGGATT 2443
QY 718 PheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThr 737
Db 2444 TACATTACCGGTAATAATTGACAAATGTTTATAGGATTTCTCCGTTCCGATGATCGCGCAAG 2503
QY 738 ThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThr 757
Db 2504 ACATGGGTGCGTATCAACGACGACGACGATATGGCGCTGTTGATATCGCAATAACA 2563
QY 758 GlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyr 777

Db 2564 GGTGACCCGAGAGTATACGACGCTGTATATATTCGCACCAACGAGAGAGTATTGTTTAC 2623
QY 778 GlyAspIleGlyGlyAlaProSerGlySerProSerProSer 791
Db 2624 GCGGAA-----CCTGCTTCAGATGAGCCTGTACCCACT 2656
RESULT 8
ABI292929 Agaricus bisporus mRNA 2579 bp mRNA linear PLN 08-AUG-2000
DEFINITION Agaricus bisporus mRNA for CEL6 protein.
ACCESSION AJ292929
VERSION AJ292929.1 GI:9755896
KEYWORDS cel6 gene.
SOURCE Agaricus bisporus
ORGANISM Agaricus bisporus
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Agaricaceae; Agaricus.
1
Morales-Almora,P. and Thurston,C.F.
Molecular analysis of the cellulolytic genes in Agaricus bisporus
Unpublished
2 (bases 1 to 2579)
Morales-Almora,P.
Direct Submission
Submitted (07-AUG-2000) Morales-Almora P., Microbiology, King's
College London, 150 Stamford Street, London, SE1 8WA, UNITED
KINGDOM
FEATURES
source
1..2579
/organism="Agaricus bisporus"
/mol_type="mRNA"
/strain="D649"
/db_xref="taxon:5341"
/clone="PMR152"
/dev_stage="mycelium"
/country="United Kingdom"
1..2579
/gene="cel6"
28..2448
/gene="cel6"
/function="putative cellulase"
/codon_start=1
/product="CEL6 protein"
/protein_id="CA02964.1"
/db_xref="GI:9755897"
/db_xref="GOA:Q9P4T8"
/translation="MPSRVKFWITLTGTFSLVYQAAQSQSVRWQNVKIGGGGFPVG
IVPNSEKGLAVARTDIGGAYKLNADDTWTPLLDPADDSRWNYWQVDALATDPVQPNR
LYLATGLTNSWDPSNGHLLISSDYGKTFDAAPLPKLGSMFGRGMRGLVVDPNLN
SLYFGARSGNLKSTNSGRSKVTSFTDTSFVDPDSDTGLNSDKLGIAWTLID
KASGSGSATPRIFYGVANKGNSIYISYNGSSWSAVAGQPTSFILPHKGVLSPSERA
LYITYSDGGPYDGTSGAVYKYSIDTAWTDITPVSGSDLFFGFGGLALDTQRNGTVM
VAALNSWPDGQIFRSTNGASWTPLNWAIVYRPLNKKYSYNAALAPWIGNIVDVTM
GNLGNWPMESLIDPDSNHLVGTGETIYGRDILLKWDASHNVTIKSLADGVEETS
VOALISPSGPPPLVSAIGDVGFGVHQSLEAPSAOFTNPKWTSADIDFAGNPSNVV
RIGTSDTSKGQVLAISSDYVTWNGHFGADPNVQGGKVAISADAILLWRTNGNGVMV
SRNQATFNIFYGASGKTFYVTDNNGKTSFSAHSLGTSATSFVFDITVHPVSGDIWASDT
KGLFHSIDTSIGTISGAIGAWGALGAPRSTGGYPAPVFAAANYGGEVAFPRSDRG
VNWVKINDAAHGFGAASACMAADPRVGRVYIGTNGRGI FYGIDVAGSAPPPTSATS
TTSVSTTTTSTRSTTTTSTRSTTTSSNGNSFVSGFGCGGCGTGTCTCTCTGTCVQCAQNP
FYSQCIPI"
28..96
/gene="cel6"
97..2445
/gene="cel6"
/product="CEL6"
sig_peptide
mat_peptide
ORIGIN
Alignment Scores: 9e-46 Length: 2579
Pred. No.: 1784.00 Matches: 355
Score:

Db 839 GCGAATGCGCTGCTTCCTTACGATGGCACAACGAAACAGTCCACACATCACC 898
 Qy 325 SerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPhe 344
 Db 899 AGCGAGGTGTGGACGGATATAAGTCCC-----ACCTCCCTTGGCGAGTACCTACTATC 949
 Qy 345 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 364
 Db 950 GGCACGGTGTCTGTCCGTTGACCTCCAGATTCTCGCACCTTGATGGTGGCTGCACATC 1009
 Qy 365 IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyValAlaThrTrpThr 384
 Db 1010 AACTCTGGTGGCGGATGAGCTCATCTTCGTAGCACTGACTCTGGAGCGACCTGGTCC 1069
 Qy 385 ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 404
 Db 1070 CCTATTGGAGTGGACGGTATATCCGAGCATCAATTACTACTACAGCTATGATATCAGC 1129
 Qy 405 AlaGluProTrpLeuThrPheGlyValGlnProAsnPro---ProValProSerProLys 423
 Db 1130 AACGCCCTTGGATCCAGACACGACTTCGACCGACCAATTCCTCCGGTT-----CGC 1180
 Qy 424 LeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr 443
 Db 1181 GTCCGCTGGATGGTTGAAGCTCTTGCCATTGACCATTTGACTCCAACCACTGGCTCTTAC 1240
 Qy 444 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGln 463
 Db 1241 GGCACGGSCCTGACAGTATATGGCGGACACGACCTGACCAATTGGGACTCGAAGCACAA 1300
 Qy 464 IleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSer 483
 Db 1301 GTGACGGTCAAGTCACTCGCGCTCGTATTGAGGAGATGGCAGTCTCGACTGATCAG 1360
 Qy 484 ProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla 503
 Db 1361 CGGCCCGCGGACCGCGCTGTGTCCTGTGGCGACGCGCGGTTTCTACCACTCG 1420
 Qy 504 AspValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSer 523
 Db 1421 GATCTGGACGACGCCAACCAACCGTACCACACGCGACCTACGCGCACGACCAACGCG 1480
 Qy 524 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 543
 Db 1481 ATCGACTACGCGCGCAACAAGCTTCCCAACATCGTCCGTCGCGAGCTAGCGAC----- 1534
 Qy 544 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 563
 Db 1535 GACTATCCGACC-----CTGGCTCTGTCTCGAATTCGATCGACTTGGTATGCA 1585
 Qy 564 GlySerGluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySer 583
 Db 1586 GACTAGCGCGCTCAACGACGACCGCGCGGCGGTGGCTCTTCCCGCATGGTGAC 1645
 Qy 584 ArgPheValTrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGly 603
 Db 1646 ACGTGCTTCTCATGTCCAGCACCTCGCGCGGCTGGTGTGCAAGTCGCGAGGCG----- 1699
 Qy 604 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 623
 Db 1700 --ACCTTGACTCGCGTTTCTCTCTCCGCTCCGGCGCGGTCTATCGCTCCGACAAAGTCC 1756
 Qy 624 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 643
 Db 1757 GACAACAGGTCTTTCTATGGTGTATCCGCGGCGCAATCTAGTGTGGAAGAACACAGCG 1816
 Qy 644 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPhe 663
 Db 1817 ACCTCTCTTCAACCAAGACCGTGAGC---CTGGGGTCCAGCACGACCGTCAACGCCATC--- 1870
 Qy 664 HisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHis 683
 Db 1871 CGCGCGCATCCCTCCATCCGCGCGATGTCTGGCGGTCCAGCCGACCAAGGCGCTCTGGCAT 1930

Qy 684 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702
 Db 1931 TCCAGGATTACGGTAGTACGTTTACGAGATCGGACGGCGTTCACGGCGGCTGGAGC 1990
 Qy 703 ValGlyPheGlySerAlaProGlySerSerTyrProAlaValPheValValGlyThr 722
 Db 1991 TTTGGGTTTGGCAGCGCTCTCCACGGGCACTAGCTCGTGTATCTACGGGTTCTTCACC 2050
 Qy 723 IleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIle 742
 Db 2051 ATCGACGGTCCGCGGGGCTGTTCAAGAGCGAGGACGCGACCACTGGCAGGTGATC 2110
 Qy 743 AsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
 Db 2111 TCGGATCGTCCACGGATTCCGCTCG---GGTCCGCCACAGCTGGTGAACGGCGACTG 2167
 Qy 761 AlaAsnLeuArgArgVal-TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
 Db 2168 CAGACTACGGCGCGGTGTTTCGTGGCACGAACGCGCGGCGCATCTTCTACGGCAAT-- 2225
 Qy 780 eGlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLe 800
 Db 2226 -----CCACGCGGAGCTCGCGGCGCCACGCGCGCGCGCGCGACGACGAC 2275
 Qy 800 uSerProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 820
 Db 2276 ACCTCCAGACCTCG-ACGACGGTATCGACGAGTGTGAACACCAACGACGAGCGGCTC 2334
 Qy 820 oSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 840
 Db 2335 GACCACTTCCAGCACCAACCGTGAAGACCACTAGTACTCGTCCAGCAGCACTTCAA 2394
 Qy 840 oSerAlaProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 860
 Db 2395 GCGCAGCTCGACCAACCAAGACTACCAACGACGACGACGACGACGACGACGACGAC 2454
 Qy 860 oSerProThrProSerSerSerProValSerGlyGly 872
 Db 2455 CACCGCGACGGCTCGGCTACGCTCAGTGTGCGGC 2491
 RESULT 10
 AY040839 3959 bp DNA linear PLN 01-APR-2002
 LOCUS Aspergillus niger endoglucanase C (eglc) gene, complete cds.
 DEFINITION AY040839
 ACCESSION AY040839.1 GI:19879405
 VERSION
 KEYWORDS
 SOURCE Aspergillus niger
 ORGANISM Aspergillus niger
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 REFERENCE 1 (bases 1 to 3959)
 AUTHORS Hasper,A.A., Dekkers,E., van Mil,M., van de Vondervoort,P.J. and de
 Graaff,L.H.
 TITLE Eglc, a new endoglucanase from Aspergillus niger with major
 activity towards xyloglucan
 JOURNAL Appl. Environ. Microbiol. 68 (4), 1556-1560 (2002)
 MEDLINE 21914059
 PUBMED 11916668
 REFERENCE 2 (bases 1 to 3959)
 AUTHORS Hasper,A.A., Dekkers,E. and de Graaff,L.H.
 Direct Submission
 TITLE Submitted (18-JUN-2001) Molecular Genetics of Industrial
 JOURNAL Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen
 6703HA, The Netherlands
 FEATURES Location/Qualifiers
 source 1..3959
 /organism="Aspergillus niger"
 /mol_type="genomic DNA"
 /db_xref="taxon:5061"
 <1038..3841
 /gene="eglc"
 gene


```

mRNA
1088 join(<1038..1137,1185..1329,1373..1571,1618..1858,
1908..1961,2007..3841)
/gene="eglic"
/product="endoglucanase C"
join(1038..1137,1185..1329,1373..1571,1618..1858,
1908..1961,2007..3841)
/gene="eglic"
/codon_start=1
/product="endoglucanase C"
/protein_id="AAK77227.1"
/db_xref="GI:19879406"
/translation="MRRFAALVAGYLLDSVHAAASQATYTKNVVTVGGGGTPIGVF
NPAGKAVAYARTDIDGGAYRLNSDDTWTPLMDWANNWNHWDGIDAIADPDVDTRVV
AVGMVNDPNDGSLRSDQDTEETKLPKVGGMPPRGVGERLAVDPNDNSIL
YFGARSGLWKSVDYGETWSNVAFKWTGTIFQDSSSTVTSDFVGLAWTFDSTSGS
SGSPTRIFVGVVDVTGSEVFSVSDAGETWTWGPMPFPLPHKGLILSPSEHTLIYSY
SNAGPDTYNTGTHVKNTISGVWMDISPTSMTDYTYGGGLAVDLQVPGTVMVAALN
CWPEDELIWSTDSGWTSPITWANGYPSINYYSYDISNAPWLQDDTDFEPPRVG
WMVEALADPDSDDHLYGTCTIYGGHDLQNDSEHNVTTIESLAVGIEMAVLGLIT
PPGCPALLSAVDDGGPYHTSLTAPSQYHTPTYSSTNGIDYAGNKPANIVRSQSD
SDPTLUSLFGESWYADYAASSSTATGQVALSADADTILLMSDGAIRSANATLSA
VSLPSGAVATSKDANNYFYAGSGSYLSSDPAATFTVTTLGSSSTANAIARAQPS
LAGDVWSTDTGLFHSNTYKGSFTQIGSGCTEGMSFGPFGPSDGDYPLVFGFTVPG
VTGLFKTEDQGVNMQIISDAEHGFGSANNVNDLQNYGRVFGVNGRGIIFYGDPGK
TLPSATATASASATKVSSTSTSKVSGSTTVSSSTATTITSSIKSTLTLTTPKS
SSSTTSSTSTATGTASAYGCGSGGTGTPQCPGWTCTYENEYSQCKSIPIGIATR
G"

Alignment Scores:
Pred. No.: 9,17e-43 Length: 3959
Score: 1690.00 Matches: 390
Percent Similarity: 53.46% Conservative: 128
Best Local Similarity: 40.25% Mismatches: 315
Query Match: 32.91% Indels: 137
DB: 8 Gaps: 26

US-09-917-376-1 (1-957) x AY040839 (1-3959)

QY 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIle 37
Db 1026 TCACCTCGTCACCATGCGCGGTACATTCGACGCCCTCGTAGCG---GGCTACCTG----- 1076

QY 38 AlaIleThrAlaSerProAlaHiAlaAlaThrThrGlnProTyrThrTrpSerAsnVal 57
Db 1077 -----CTGACTCGGTCCCGCTCCCGCATCGCGGCATACCTGGAAGAATGTG 1127

QY 58 AlaIle-----Gly 60
Db 1128 GTCAC-AGGGGGTATGACATATCTTGGCAATATGAGGGATGACAAAGCTAATGTATGACGC 1186

QY 61 GlyGlyGlyPheValAspGlyLeValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
Db 1187 GCGGAGGCTTCACGCGGGTATCGTCTTCAACCCCTTCAGCTAAAGTGTGCGCATATGCT 1246

QY 81 ArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProIle 100
Db 1247 CGCACCGACATCGCGGGTGCATATCGGTGAAC---TCGACGATACATGACCCCGTTG 1303

QY 101 LeuAspTrpValGlyTrpAsnAsnTrp-----GlyTyr-Asn----- 112
Db 1304 ATGACATGGGCCAACAACTAACTGGTACGTAGCTTCTTACGAGATGCCTAGGCTC 1363

QY 113 -----GlyValValSerIleAlaAlaAspProIleAsnThrAsnIly 126
Db 1364 ACTGTCACGACCGATTGGGGCATATGATCGCATCGGACCGACCCGTCGATACCGATCG 1423

QY 126 sValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLe 146
Db 1424 GGTATATGTCAGTTGGCATGTACACCAACGACTGGGATCCCAATGACGCGCTCCATCCT 1483

QY 146 uArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGl 166
Db 146 uArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGl 166

```

```

Db 1484 CGGCTCGACCGACCGAGCGGATACCTGGGAAGAGACAAAGCTTCCTTCAAAGTTGGAGG 1543
QY 166 yAsnMetProGlyArgGlyMetGlyGlu----- 175
Db 1544 GAATATGCCCGGTTCGGGTGTGGCGAGGTAAAGTCTATATCTGGGTTTGTAGCATAC 1603
QY 176 -----ArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyA 191
Db 1604 ACTGACCGCTCGACGCTCTGGCTGTGGATCCGAATGACAAACAGCATCTTGTTACTTTGGTG 1663
QY 191 laProSerGlyGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetTr 211
Db 1664 CCGGAGTGCATAGCTTGTGGAAGACACGAGCATAGCGCGGAGCATGCTCCAATGTGA 1723
QY 211 hrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 231
Db 1724 CTGCCTTCAAGTGGACCGGCATTTATTTCCAAAGATTCCAGCTCGACCTGACCTGCGTCTC 1777
QY 231 erAspIleGlnGlyValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnA 251
Db 1778 CTGATCCGGTGGGAATTGCTGGGTCACTTCGATTCGACCTCAGGTCAGGTTCAGTGGCTCTC 1837
QY 251 laSerLysThrIlePheVal----- 257
Db 1838 CGACTCCGCGCATCTTTGTGCTATGTGAAATAATCTAGATATCTCAGGACACGCCCATCTG 1897
QY 258 -----GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGly--- 273
Db 1898 ACCATTACAGGTGTGGTGGACACTGGAGATCCGCTTTGTGTCGAAAGATGTCGGCGAA 1957
QY 274 -----AlaThrTrpGln 277
Db 1958 ACTTGTAAGCTTCTCCACGATCGCAATTAACCCATCTTAACCACTTAGGACTTGG--- 2014
QY 278 AlaValProGlyAlaProThr---GlyPheIleProHisLysGlyValPheAspProVal 296
Db 2015 ---GTATCTGGAGAGCCCATGTACGGCTTCTCTCCCAATAAAGGCATCTCTGTCGCTCT 2071
QY 297 AsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGly 316
Db 2072 GAGCACACACTCTACATTTCTACTTCCAAACGCGCGGTCCATACGATGGCAACCAATGCG 2131
QY 317 AspValTrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSer 336
Db 2132 ACCGTCCACAGTACACATCACACGCGCGGTCTGACGCGCATCTAGTCCACCTCCATG 2191
QY 337 ThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro 356
Db 2192 ACAGACACC-----TACTATGGCTATGCGGACTGGCAGTCGATCTCCAAGTGCCC 2242
QY 357 AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSer 376
Db 2243 GGAACGTGTCATGTTAGCAGCATCAACTGCTGTGGCGCTGATGAGCTGATCTGGCGCAGC 2302
QY 377 ThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSer 396
Db 2303 ACGGATTCTGGGCGCACCTGGTCAACCATCTGGGCTTGAACGGGTATCCAGCATCAAC 2362
QY 397 LeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsn 416
Db 2363 TACTACTACTCTTATGACATCTCCAAACGCCCTTACAGCAGCAGCATCTACAGAC 2422
QY 417 Pro----ProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPro 435
Db 2423 GAATTTCCGCT-----CGCGTGGCTGGATGGTCCAAAGCCCTCGCCATTGATGCC 2473
QY 436 PheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeu 455
Db 2474 TTCGACTCTGACCTTACGGCTTTACGGCATCGCGAGACCATCTACCGCGGCCACCTC 2533
QY 456 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 475
Db 2534 CAAATTTGGGACGACGACACCAACGTAATCTATCGATCTCTTGGCCGTAGGATCGAGAA 2593

```



```

476 ThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGly 495
2594 ATGGCTGTTCGGGTCTTCTATCACTCCCGAGCGGCGCTTCCTCTCTCGGTGCGC 2653
496 AspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSer 515
2654 GACGATGGGGCTTCTACCACTTCCCTGACCAACCGCCCATCAATACTACCAACACC 2713
516 ProValPheThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleVal 535
2714 CGGACTTACAGTTCACAAACCGCATCTGACTACGCGCGCAACAGCCGCTAACATCGTC 2773
536 ArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThr 555
2774 CGCTCTGGATCCAGCGAT-----TCGACCCCTACTCTGCCTTTATCCTCC 2818
556 AspGlyGlyLyAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr 575
2819 AGTTTCGGCGAATCTTGGTACCGGACTAGCTGCTTCTTCTCTACTCGGACCGCCAA 2878
576 ValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnPro 595
2879 GTGCGCCTCTACGCGATCGACACCACTCTCTCTCATCAACAGCGACGGT----- 2929
596 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsn 615
2930 ---GCCTATCTGTCGCAAACTCGGCTACTCTTAGCGCGACTCTCTCTGCGCCTCGA 2986
616 AlaGlnIleArgSerAspArgValAsnProLyThrPheTyrAlaLeuSerAsnGlyThr 635
2987 GCTGTGATCGCTTCCGACAAAGCTTACAACTCTTCTTAGGACCTTCGGCTCTTTC 3046
636 PheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 655
3047 TTTCTATCTTCTACCGACCGCGCGACTTC---ACTGTACACCAACCCCTCGCGAGT 3103
656 SerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 675
3104 AGCACAACCGCAACGCCATC---CGCGCGCAACCTCTCTCGCAGCGATGTCGGTCT 3160
676 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 694
3161 AGTACGACACAGAGTCTTCTCCACTCAACCACTACGCGAAATCTTCCACCCAGATGGG 3220
695 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 714
3221 AGCGGATGCACAGCGGTGGAGTTCGGATTCGGAAGCCCTCTTCGACGCGGACTAC 3280
715 ProAlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAsp 734
3281 CGGTCTTATTTGGATTCTTTACCGTCGATGGTGTGTACCGGTCTCTTCAAAACCGAAGAT 3340
735 CysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsn---TrpGly 753
3341 CAGGGTGTAAATGGCAGATAAATTCGACCGCGCAGCAGCGGTTCGGGTCTGTCGCGCA 3400
754 GlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArg 773
3401 AATGTAGTCAATGGGATCTTCAGACTATGCGAGGGGTGTGTGTGGCAATGGCGGA 3460
774 GlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSerValSer 793
3461 GGAATCTTCTATGTGAT-----CCGAGTGGAACT---CTGCGGAGTGGCGACT 3505
794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSerProSer 813
3506 GCTACTCGCTCTAGTGTCTGTCGACAGCGGTGAAGAGTTTCGACCTCGACTTCGACATCC 3565
814 ProSerProSerSerSerProSerProSerProSerProSerProSerProSerProSer 833
3566 AAGTAGGTTCCAGCACTGTTTCTTCGAGTACGCGCTAACTACTACTACGAGCAGT 3625
834 ProSerArgSerProSerProSerProSerProSerProSerProSerProSerProSer 853
3626 ATTAAGTCGACCACTTTGACGACGACTACAAAGTCGAGTAGTTCGACTACTTCGACGCA 3685
854 SerSerProSerSerSerProSer-----ProThr 863
3686 TCCTACTGCTACTGGGACGCGCTCGCATATGGCAGTCGGTGGGAGCGGGTTTACGGGA 3745
864 ProSerSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAsn----- 880
3746 CCGAGCAGTGGCGG-----AGTGGTGCACGCTGCACGATGAGATGATATTATTCG 3799
881 -----AspSerAlaProGly 885
3800 CAGTGTAAAGTCTATCCTCTGCT 3820
AK110506 2724 bp mRNA linear PLN 24-JUL-2003
Oryza sativa (japonica cultivar-group) cDNA clone:002-167-F04, full
insert sequence.
AK110506
AK110506.1 GI:32995715
FLI_CDNA; oligo capping.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
SOURCE
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1
REFERENCE
AUTHORS
The Rice Full-Length cDNA Consortium, National Institute of
Agronomical Sciences Rice Full-Length cDNA Project Team,
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Ootomo,Y., Murakami,K.,
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
Narioka,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN,
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
Hara,A., Hashidume,M., Hayatsu,N., Imotani,K., Iehi,Y., Itoh,M.,
Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Yoshino,M. and Hayashizaki,Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
JOURNAL
MEDLINE 301 (5631), 376-379 (2003)
PUBMED 22752273
12869764
REFERENCE
AUTHORS
2 (bases 1 to 2724)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hirooka,T.,
Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Kagawa,I.,
Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kawai,I.,
Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
Namiki,T., Narioka,R., Niikura,J., Nishi,K., Nomura,K.,
Numaaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
Osato,N., Ota,Y., Ootomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Fukuda,Y.,
Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,T.,
Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
Yoshino,M.
```


Yoshimura, A.
 Direct Submission
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail: shikuchienias.affrc.go.jp,
 Tel:81-29-838-7007, Fax:81-29-838-7007)
 This clone is one of the 29k full-length cDNA clones from japonica
 rice.
 URL : http://cdna01.dna.affrc.go.jp/cDNA/
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and
 Yamamoto, W.

FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
 Kodama, T., Kurotsaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
 Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
 Yoshimura, A., Matsubara, K. and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
 Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
 Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
 Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
 Yasunishi, A. and Hayashizaki, Y.

FEATURES
 Location/Qualifiers
 1..2724
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultiivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="002-167-F04"

ORIGIN

Alignment Scores:
 Pred. No.: 3,36e-41 Length: 2724
 Score: 1633.00 Matches: 365
 Percent Similarity: 58.11% Conservative: 151
 Best Local Similarity: 41.10% Mismatches: 307
 Query Match: 31.80% Indels: 65
 DB: 8 Gaps: 27

US-09-917-376-1 (1-957) x AK110506 (1-2724)

QY 37 IleAlaIleThrAlaSerProLahisAlaAlaThrThrGlnProTyrThrTrpSerAen 56
 DB 88 ATCTCGATAACCGCTCTCCG-----GCCGTGGTCGAGCGGCTCTACTCTCGAAGAT 141
 QY 57 ValAlaIle---GlyGlyGlyGlyPheValAlaPheGlyIleValPheAsnGluGlyAlaPro 75
 DB 142 GCCAAGTTGGAGCGCGGTGGCTTTGTGCGCGGCATTTGATTCATCTCCTACCCAGGT 201
 QY 76 GlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGly 95
 DB 202 GCGGTGCTTATGCTCGAACGGATATAGAGGCGCTGTACAGTGGAAC---AGTGATGAC 258
 QY 96 ArgTrpIleProLeuLeuAspTrpVal-----GlyTrpAsnAsnTrpGlyTyr 111
 DB 259 TCGTGGAGCCCATCACCAGCTCCATCACCAGCAATGCGAATGGCAAAATTGG----- 312
 QY 112 AenglyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131
 DB 313 ---GGAGTGATGCGCTTGTCTTATCCCGAGGAGCCCAATGTGTGTGCTTCGCGCGG 369

QY 132 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 151
 DB 370 GGTGTGTAACCAACTCTCTGGATCTAGCCCGCGGCATCTACAAGAGCAGGACAAG 429
 QY 152 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 171
 DB 430 GGTGCGACTTGGACATCTTCAACGCTCAGCTTCAAGGTGCGTGTAAACATGCTGCGCCG 489
 QY 172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191
 DB 490 GGCATGGCGCAGCGCTGCTGCATCCCAAGAACTCCAAATTTGTGTTTCGCGCGCA 549
 QY 192 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211
 DB 550 CGCTCTGGAACCGGCTCTGGAGAGAGACCGAGCGCGGCAAGACCTTCAAGAGGTGCG 609
 QY 212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 231
 DB 610 TCTTCAAGGCTGTGTGGTACCTTACATCCGGTCCCAGCGCAGCAGCAGCGCTTACAACAGC 669
 QY 232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 251
 DB 670 GACATCAACGCGCTGACCTTTGTGACCTTTGTACTCGCCTCCGCAACCATCAATGTTGCC 729
 QY 252 SerLysThrIlePheValGlyValAlaAsp---ProAsnAsnProValPheTrpSerArg 270
 DB 730 ACTGCGCGCATATTCGTAGGACCCCGCATACGTCACGGCTCGCTTACGTTTCTGAG 789
 QY 271 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 290
 DB 790 GATGCGCGCAGCAGCATGGACCGGTCAAGGGCCAGCGCTGGCAATACTTCTCTCATAG 849
 QY 291 GlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 310
 DB 850 TGCAGGCTGCGAGCCCGACCGAGCGCACTCTTACGCGCAGCAGCAACGGGCCCT 909
 QY 311 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpTrpArg 330
 DB 910 TACGATGTGGCTCTGCGGCTCTGTGGCGATACGGCTCGCGCAGCGCAGCGGACGAG 969
 QY 331 IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 350
 DB 970 ATCATCTCTGCTCCGTTAGCGACCTGACC-----TACGGATTCGCGCGCTCTCC 1020
 QY 351 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 370
 DB 1021 GTGGACATGAAGAAGCCAGAACTCTGGTCGTGCGCAGTCTCAACTCTTTGGTGGCTGAT 1080
 QY 371 ThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspTrpThr 390
 DB 1081 GCGCAGATCTCCGCGACCGACCTCTGGCAAGAGCGTGTCAAGGCTCTGGCAGTGGGT 1140
 QY 391 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr 410
 DB 1141 AACTATCCGAGATGATCTACTATTACGGCATCTCGACTCCCAAGGCCCTCTGG----- 1194
 QY 411 PheGlyValGlnProAsn---ProProValProSerProLysLeuGlyTrpMetAspGlu 429
 DB 1195 -----ATCCGCAAGAACTTCTCTCGATACCGACTCCAAGTGGCTGGCTGATGATCGAG 1248
 QY 430 AlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeu 449
 DB 1249 TCTTTCGAGATCAACCCGACTGACAGCGACCACTTCTCTGTACGGCACCGGACCTGAC 1308
 QY 450 TyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMet 469
 DB 1309 TACGGTGGCCAGCCTGACCACTGGGACCAACCAAGCGCAACATCACTTCAGTCCCTG 1368
 QY 470 ValLysGlyLeuGluThrAlaValAlaAsnAspLeuIleSerProProSerGlyAlaPro 489
 DB 1369 GCCGACGGCATCGAGAGACCGCTGTTCAGGAGGTGGCTCTCGGCCCGCGCGCAGGAG 1428

Alignment Scores:

Alignment Scores:			
Pred. No.:	3.46e-41	Length:	2517
Score:	1631.50	Matches:	353
Percent Similarity:	75.23%	Conservative:	130
Best Local Similarity:	41.82%	Mismatches:	297
Query Match:	31.77%	Indels:	63
DB:	8	Gaps:	23

US-09-917-376-1 (1-957) x AY281371 (1-2517)

Qy	25	SerpheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla	44
Db	10	TCCTGAGTCTTGCCCTTGCTGGGGCCGTC-----ATCCCTGC	51
Qy	45	HisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGly	63
Db	52	CATGCTGCC-----TTTTTCATGGAGAACGTCGAAGTCGCGCGCGCGCGCGC	99
Qy	64	PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp	83
Db	100	TTTGTCGCCCGGCAATCATCTTCATCCCMAGCAAAAGCGGTAGCATATGCACGACAGAT	159
Qy	84	IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp	103
Db	160	ATTGGCGGGCTGTACCGCTCAAC---GCCGACGACTCATGGACCGCGGTACGATGGG	216
Qy	104	Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla	119
Db	217	ATTGCTGATAATGCCGGCTGGCACAACTGG-----GGCATCGACGCTGTGCGCT	267
Qy	120	AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp	139
Db	268	GATCCGCGAGGACGATCAAAAGAGTGATGCCGATCGGCATGATATACGAACAGCTGGAT	327
Qy	140	ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro	159
Db	328	CCGAGTATGAGCCATCATTCGCTCGTCAGACCGCGCGGCACAGTGTCTTCACCAAC	387
Qy	160	LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal	179
Db	388	TTGCGCTTCAAAGTCCGGGGTAACATGCCAGGACGCGGACCGGAGAGCGCTCTGGCTGC	447
Qy	180	AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArg	199
Db	448	GATCCGGCCAACTCCAAACATCATCTACTCTTTGGTCTCGCTCAGSAAACGGCCCTCTGGAAG	507
Qy	200	SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyr	219
Db	508	TCTACGGACGGCGGGTGACCTTTTCCAAAGGTCTGCTGTTCCAGGCACCTGGGACGTAC	567
Qy	220	IleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal	239
Db	568	ATCCCGACCCGAGTGATTCCAAACGGCTACAAACGCGCAAGCAAGCAAGGACTCATGTGGT	627
Qy	240	AlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal	259
Db	628	ACGTTTCGACTCAACACGACGACGACCGCGGGGAGCCACGCTCGTATCTTTGTTGGCAGC	687
Qy	260	AlaAspPro---AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla	278
Db	688	GCTGATTAACATCACTCTTCAGTCTATGTGAGCAGCAATCCCGGCTCCACGTGGAGTGCT	747
Qy	279	ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis	298
Db	748	GTACCGGGGACCCAGGGGAATACCTTCTCACAAAGGGGAAACCTGCAGCCACGACAGAGAAG	807
Qy	299	ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal	318
Db	808	GCCTTGATCTGACCTATTCCGATGGCACAGGGCCGTATGATGSCACACTTGGCTCAGTG	867
Qy	319	TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp	338

Db	160952	CTGCTCAAGTCGACCGACCGGGCGCCACCTTGGGCGAGCGCGCGCTTCCCG	160899	Qy	562	---PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla	580
Qy	217	GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal	236	Db	159965	CCCTTCGAGGCCACCCGACATCGCCACGAGCGACCGGACCGATCGCCACGAGTCC	159906
Db	160898	-----CGGAAGGCGAACTCTCTCGGCGAGGAGTC	160869	Qy	581	AspGlySerArgPheValTyrPala	596
Qy	237	ValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePhe	256	Db	159905	GACGCGCGCACACTGCTGTGTCCTTCTGTGTCATCGGACGCGC	159861
Db	160868	GTGTTCT	160836	Qy	597	ValTyrAla	612
Qy	257	ValGlyValAlaAspProAsnAsnPro	271	Db	159860	ACGTACGCGGCCACCGCTCGACGACAAACGCGCGAGCTGGTCCGAGGTCTCTCTCTTC	159801
Db	160835	GCCGCTGGGTGACGCGGACCGCACCTCGGCGACGGCGAACCCTGTACCGCACGCGCGAC	160776	Qy	613	ProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla	630
Qy	272	GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe	288	Db	159800	CCGAAGGCGGCACACCGGTTCGCCGACCCGCGCATCCGACGCGCTTCTACGCTACGAC	159741
Db	160775	--GGCAGCACTGGGGGCGCTCCCGCGCGCCCTCCGCGCACCTCCGCGCAAGGTCCCG	160719	Qy	631	LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAla	650
Qy	289	HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly	308	Db	159740	TTCGACAATGGAACGCTATACGCCAGCACTGACAGTGGCGGTTCGTTTCACGCGCGCGTGC	159681
Db	160718	CTCCGCGCGCGTACGACACGACACCCCGCGAGCTGTACGTGACGTACGCGGACGACCC	160659	Qy	651	AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlu	670
Qy	309	GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp	328	Db	159680	GGCGGACTGCC--TCCGCGCACAGCCAGTTTCAAGCTGGTTCGCGCGCGCGGACGAAAGC	159624
Db	160658	GCCTCGGCGCGCAGTCCGACGCGAGCTGCACAGCTGCGTACCGCCACGCGGAGCTGG	160599	Qy	671	GlyAspLeuTrpLeuAlaAlaSer	689
Qy	329	ThrArgIleSerProVal	344	Db	159623	GCGGACCTGTGTGCTCTCCGCCAAATGAAACGCGGTCTACCGTCCACCGACGCGGGGAC	159564
Db	160598	ACCGAGGTACCCCGGTGAAGCGCGGCGGACGACGAGCGGCTCGCGGACACCTTTC	160539	Qy	690	SerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla	709
Qy	345	GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln	364	Db	159563	ACCTTCGCCAGGATCGACAGCTGCTGGGCTCTCGTACACCTCGGTTTCGCAAGCGGCC	159504
Db	160538	GCCTACGCGCGGTGCGCGTGCACGCCCGCGCGCGCGCGCGCTCTCTCCACCAAC	160479	Qy	710	ProGlySerSerTyrProAlaValPheValGlyThrIleGlyGlyValThrGlyAla	729
Qy	365	IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThr	384	Db	159503	GACGCGCGCGCATACCGCGCATCTACAGGTGGCTCGACGCGGACCATCACCGCGCTC	159444
Db	160478	AACCGCTGGCGGACGCGGACACGCTCTTCGCGTCCAGGACGCGCGCGCTACCTGGAGC	160419	Qy	730	TyrArgSerAspAspCysGlyThrThrTrpValIleLeuAsnAspAspGlnHisGlnTyr	749
Qy	385	ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer	404	Db	159443	TACCGTCCGACGACGCCCGCGAGCATGGTCCGATCAACGACGACGCGCCACCATGG	159384
Db	160418	TCCTCAGGAC	160386	Qy	750	GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGly	769
Qy	405	AlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeu	424	Db	159383	GGGTGATCGGCGAGCGCGTCTCGTCCGTCACCGCGCATCCACGCGCGGTCTACTCGCC	159324
Db	160385	GAGACTCCCTTCTCGACTGGGCGGACGAC	160344	Qy	770	ThrAsnGlyArgGlyIleValTyrGlyAspIle	784
Qy	425	GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly	444	Db	159323	ACCAACGCGCGCGCATCAGTACGCGGAGCGGTCTGATGCCCGGCTCACGCGATGCCA	159264
Db	160343	GCGTGGTGGATCCAGCGGCTCGCGTCCGACCGTACCGTCCGACGACGTCGTGTACGGG	160284	Qy	785	SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro	804
Qy	445	ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIle	464	Db	159263	CCCGGGCGCGCATCTCTTGTGGTGGCG--ACTACAAACCCGAGCATGGCCGCGGAGG	159207
Db	160283	ACCGGCGGACCTCTACGGCACCCCGGACCTCAAGCGTGG	160242	Qy	805	Ser	812
Qy	465	HisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPro	484	Db	159206	TCTGGCAGGAGGACGTACGCTCATGAGGAGCGCGCGTCACTCGTCCGTCGCGCG	159147
Db	160241	-----GCACCGCGATCCGCGCTCGAGGAGCGCGCTGCGCCAACTGATCTCGCC	160188	Qy	813	SerProSerProSerSerPro	820
Qy	485	ProSerGly	503	Db	159146	TCTTCTCTGGGCGGAGCTCGAACCCACCCCGCGGCGGAGGAGTTCGGTTCGCTCGAC	159087
Db	160187	CGGTGCGGAGGACACCTGATACGCGCATCGGGGACATCGGTGTGATGATACACGAG	160128	Qy	821	-----SerSerProSerProSerPro	828
Qy	504	AspValThrAlaValProSerThr	522	Db	159086	GGCTCATGGACCTGATGTCACAGAACGCGCATCGCGTCTGCTGGCCACCCCGCGCT	159027
Db	160127	CGGCTCACGCGCTCTCCGTCGGCGCATGGCGAAGAACCCCGGTTCGGGTTCGGCGACG	160068	Qy	829	SerProSerPro	832
Qy	523	SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro	542	Db	159026	CGCCCCCGCCCTGGATGGGCGCGCTGTCACCCCGAGACCTGCCCGCGACGAGGACGCGC	158967
Db	160067	GGACTCGCGCAGCGCGCGCCAGCGCGGTATGCTCGCGACGCGCTGGGCGGACAC	160008	Qy	833	-----SerProSerArgSerProSerProSerAla	842
Qy	543	SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrp	561	Db	158966	GCATCGAGTGGTGGGGCGCGCGACCACTTCGCCCACTCCAGTGGCTTACCGCGCGCT	158907
Db	160007	GGCAACGGC	159966				

[illegible]


```
gene
CDS
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM36636.1"
/db_xref="GI:21107970"
/transl_table=11
/product="D-xylulokinase"
/protein_id="AAM36639.1"
/db_xref="GI:21107973"
/translation="MSLYVGLDVGTSVKLVAYDPODRAVRGIVSGVQHQHGFVPAADGSVTPVKLWCDTS
BQQAQWIDGIVHCFAQLDGEORARVGISSVQHQHGFVPAADGSVTPVKLWCDTS

Alignment Scores:
Pred. No.: 7,54e-27 Length: 14520
Score: 1179.50 Matches: 312
Percent Similarity: 43.94% Conservative: 134
Best Local Similarity: 30.74% Mismatches: 357
Query Match: 22.97% Indels: 213
DB: 1 Gaps: 29

US-09-917-376-1 (1-957) x AB011809 (1-14520)
QY 27 AlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46
Db 150 GCGGCTGCTGCTGACGCTGGTGGTCTCTTCATGACGTCGAGGACATCCGCGGACGAGCCC 209
QY 47 AlaThrThrGlnProTyrThrTyrSerAsnValAlaIleGlyGlyGlyPheValAsp 66
Db 210 GGTACGCCACAGCCGATCAGTGGCAGCGTGGCCATCGGGGGGGGGCGGCTTTGTCAACC 269
QY 67 GlyIleValPheAsnGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86
Db 270 GGTGTGCTGTTTCATCCCGCCGACCGATCTTGCCTATGCACGACCGATGTCGGTGGT 329
QY 87 MetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyrValGlyTyr 106
Db 330 GCCTATCGTGGGACGCGGACGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 389
QY 107 AsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126
Db 390 GACGATTGGAATCTGATGGGAATCGACGCTTTTGGCGTGCATCCTCGCGACCCGACGCGC 449
QY 127 ValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspProAsnGlyValAlaIleLeu 146
Db 450 TTGTACTGGCTGCTGGCCTACATGACGAGAACGCGCGCGCAAT---GCCGCGGTGTG 506
QY 147 ArgSerSerAspGlnGlyAlaThrTyrGlnIleThrProLeuProPheLeuGlyGly 166
Db 507 CGTTTCGTCGATCGCGTGAAGCAGTTCGACGCGCGACCTGCGTTCAGCTCGGTGGC 566
QY 167 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186
Db 567 NATCAATTGGCGCGCGCAATGCGGACGCGCTGGCGCTTACCGCGACGACGCGCGGTGTG 626
QY 187 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206
Db 627 CTGTCTGCTGGCTGCTGATGTCG---GGCTGTGGCGGACGACGATCGCGCGCGCGCAC 683
QY 207 TrrSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226
Db 684 TGGCGCGGTGGAAGATTCTCTGCCGATGTCATGTCGCGCGCGCGCGCGCAACCAT 743
QY 227 ThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSer 246
Db 744 GTCGGAAGCGAACAGCGCGGTG---GGCATCGCTTCTGCTGGTGTTCGACGCGCGCGCGC 800
QY 247 SerLeuGlyGlnIleAsnLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266
Db 801 CATGCTGGCGCTCGCACACACCATCTATGTGGCGGCTTTCGACTGCGCGCAACCGCGCTG 860
QY 267 PheTyrSerArgAspGlyGlyAlaThrTyrGlnAlaValProGlyAlaProThrGlyPhe 286
Db 861 TATGTCTCCGAGATGCGGCGGCGCACCTGTGCGCGCGCTGGCGCGCGCAACCGCGCGCTG 920
```

```
gene
CDS
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM36636.1"
/db_xref="GI:21107970"
/transl_table=11
/product="D-xylulokinase"
/protein_id="AAM36639.1"
/db_xref="GI:21107973"
/translation="MSLYVGLDVGTSVKLVAYDPODRAVRGIVSGVQHQHGFVPAADGSVTPVKLWCDTS
BQQAQWIDGIVHCFAQLDGEORARVGISSVQHQHGFVPAADGSVTPVKLWCDTS

Alignment Scores:
Pred. No.: 7,54e-27 Length: 14520
Score: 1179.50 Matches: 312
Percent Similarity: 43.94% Conservative: 134
Best Local Similarity: 30.74% Mismatches: 357
Query Match: 22.97% Indels: 213
DB: 1 Gaps: 29

US-09-917-376-1 (1-957) x AB011809 (1-14520)
QY 27 AlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46
Db 150 GCGGCTGCTGCTGACGCTGGTGGTCTCTTCATGACGTCGAGGACATCCGCGGACGAGCCC 209
QY 47 AlaThrThrGlnProTyrThrTyrSerAsnValAlaIleGlyGlyGlyPheValAsp 66
Db 210 GGTACGCCACAGCCGATCAGTGGCAGCGTGGCCATCGGGGGGGGGCGGCTTTGTCAACC 269
QY 67 GlyIleValPheAsnGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86
Db 270 GGTGTGCTGTTTCATCCCGCCGACCGATCTTGCCTATGCACGACCGATGTCGGTGGT 329
QY 87 MetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyrValGlyTyr 106
Db 330 GCCTATCGTGGGACGCGGACGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 389
QY 107 AsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126
Db 390 GACGATTGGAATCTGATGGGAATCGACGCTTTTGGCGTGCATCCTCGCGACCCGACGCGC 449
QY 127 ValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspProAsnGlyValAlaIleLeu 146
Db 450 TTGTACTGGCTGCTGGCCTACATGACGAGAACGCGCGCGCAAT---GCCGCGGTGTG 506
QY 147 ArgSerSerAspGlnGlyAlaThrTyrGlnIleThrProLeuProPheLeuGlyGly 166
Db 507 CGTTTCGTCGATCGCGTGAAGCAGTTCGACGCGCGACCTGCGTTCAGCTCGGTGGC 566
QY 167 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186
Db 567 NATCAATTGGCGCGCGCAATGCGGACGCGCTGGCGCTTACCGCGACGACGCGCGGTGTG 626
QY 187 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 206
Db 627 CTGTCTGCTGGCTGCTGATGTCG---GGCTGTGGCGGACGACGATCGCGCGCGCGCAC 683
QY 207 TrrSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 226
Db 684 TGGCGCGGTGGAAGATTCTCTGCCGATGTCATGTCGCGCGCGCGCGCGCAACCAT 743
QY 227 ThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSer 246
Db 744 GTCGGAAGCGAACAGCGCGGTG---GGCATCGCTTCTGCTGGTGTTCGACGCGCGCGCGC 800
QY 247 SerLeuGlyGlnIleAsnLysThrIlePheValGlyValAlaAlaAspProAsnAsnProVal 266
Db 801 CATGCTGGCGCTCGCACACACCATCTATGTGGCGGCTTTCGACTGCGCGCAACCGCGCTG 860
QY 267 PheTyrSerArgAspGlyGlyAlaThrTyrGlnAlaValProGlyAlaProThrGlyPhe 286
Db 861 TATGTCTCCGAGATGCGGCGGCGCACCTGTGCGCGCGCTGGCGCGCGCAACCGCGCGCTG 920
```


/gene="XCC1751"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="Tomb-dependent receptor"
/protein_id="AA041042.1"
/db_xref="GI:21112844"
/translation="MMVGTLPYKKVEGFANAVNEVYDQVWQITRPVNGDACKIRG
AELGYTFDFPLPGLWGLGNTYIVYDSEAPSTADITNGQSLTVPLEGLSKNSYN
AILEYETARFCGRVAYNMRSDMLVTAGNGTGNLPVYKNGFGQDLASLRFNINDVWSI
SLDGNLLDTRRESYLGTSRYDRFVINDRYGLTLASL"
7567..9822
/gene="XCC1752"
/db_xref="XCC1752"
7567..9822
/gene="XCC1753"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="AA041043.1"
/db_xref="GI:21112845"
/translation="WMIMIDPPTKRRITTTGAWLAVLSLLLLPSTPNVRAAEPATSG
PYORSVAIGGGFVGLPFAERGLAYARTDVGAVRMDAQAOQWLTALTDMLGADD
WNLGIDAFVADPADADALYLAAGTYMHERAGNAVLRSFNRGTFERADLPFLKGN
QLRANGRLAVPHDGRVLLGSRDAGLWRSDDGRHWKAVSFDDALAGATARNH
VGRQAVFVFDVADSGNTGTPTPIYGVSTEQTSLYVSEDAGSRWAPVAGPRG
LRPMAGSGDGHVLSYGDQPDLMAGALKWFTPAQWRBISIPPOASGDGFG
WGVADVPOOPQVLLASTFRRTTDELRYSDVGGKHWAPLLADAVFDSAPATAHA
TPHMGALIDPPGNHALFVTGYIMASRNLDQFAAPQRPLOWFODRGLEETVPLD
LLSPMAGNLLSALGDGDFRDELDRALQYAGPRLTNGESIDAAGAPQWVVRSGT
VRDRNNELSGDQKQWTFASEPPAGGAGSIAIGADAAQVWVAPERGNGWT
SDFGQWRQDGLDPTAVMADRVDARWYADVASQQLYESTDAARSFRATGVQVGS
PARBRTYRQLRPPDPVAGVVLASPGKGVNRWDQTLQVLSQDPEARSILGIGKALRA
GAPALYLAGRVQGVGDFRSDDGVQWQORINDDAHRFRPVSVTGDPRIAGRVVFA
GGRGIFYGDPB"
9926..11839
/gene="XCC1753"
9926..11839
/gene="XCC1753"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="sialic acid-specific 9-O-acetyl esterase"
/protein_id="AA041044.1"
/db_xref="GI:21112846"
/translation="WTFSHISRACILGLVLSGSSNAQPSLPLIFADGAVLQREAMP
VWGATPGASITVSDGQAKTAKDAQOQWVLLPAHTAGGPPYVLSVHGDGGLQVRD
VLVDVWLASGNSNWPQLAQSDDQAVAAANDALRHFVKPSWSVOPETKLTGGD
WKAAMPENAGAFSAGVYFAKELRSTGPTIGVISTWGSALRAMDAASLGLKAEQ
NQAIEATKORDAAQAATGKRIGRWPKVDGTPQWRTPEFDRDWDSDIPVTOOWELS
GYEGWDGIAWYFTLTAAEKAGVOLGVQIDSDITYNGQVQVGTOKOMNLPV
YAVPAALRAGVQIAVRVEDLSGGGHHGDEQRFVQIGSTQRLAGKFRPAQVRV
SLTNKQPLLLNQMHPLOPPFVKGVIWYQGETNATDGVAKYRQDFAAMIQQWR
AERKARLPLFVLQANPKAGDGTDLSPWALLRESQSKTLALPATGOAVIDIGNPT
DIHPTNKDGVHRLAARHAYGETLVYSAPFARASVADQASVQDLOGSALQVR
GGTTLQGFVAGADKRFHPAQAIRIAGDRVIVRSDAVAPVAVRYGWSNPADANLVR
EQLVSPFRTDTW"

gene
CDS

Alignment Scores:
Pred. No.: 1,17e-26 Length: 11910
Score: 1170.50 Matches: 299
Percent Similarity: 48.61% Conservative: 138
Best Local Similarity: 33.26% Mismatches: 350
Query Match: 22.79% Indels: 113
DB: 1 Gaps: 28

US-09-917-376-1 (1-957) x AE012276 (1-11910)

23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42

Db	7606	ACCACACAGGAGCATGGCTGCTGTAGCCCTGCTG---CTGTGCTCTTTTCGACG	7662
Qy	43	Pro-----AlaHisAlaAlaThrGlnProTyrThrTrpSerAsnValAla	58
Db	7663	CCCAACGTCGTCGGCGCGAGCCGCCACGTCGGGCGCTACACAGTGGCGAGTGTGCC	7722
Qy	59	IleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu	78
Db	7723	ATTGGCGTGGCGGATTTGTACCGTGTGTCTGTTTCATCCCGCGAAGCTGGTCTGGCC	7782
Qy	79	TyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle	98
Db	7783	TATGCGGCACCGATGTGGTGGCGGTACCGTGGATGCGAGCGCGCAGCAGTGGACC	7842
Qy	99	ProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla	118
Db	7843	CGCTGACCGCATGCTGGCGCGCCAGCAGCTGGAACCTGATGGCATCGACGCAATTCGCC	7902
Qy	119	AlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrp	138
Db	7903	GTCGACCCCGCATGCGCATGCGCTGTATCTGGCGCGCGGACCTATATGACGACGACGC	7962
Qy	139	AspProAsnAspGlyValIleLeuArgSerSerAspGlnGlyAlaAlaThrTrpGlnIleThr	158
Db	7963	GCCGCGCAAT---GCCGCGATGCTGCTGCTTCAACCGCGCGCGCAGCTTCGAGCGCGCC	8019
Qy	159	ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla	178
Db	8020	GACCTGCGCTTCAAACTGGTGAACAGCTGGCGCGCGCAATGTGCGAGCGCTGGCG	8079
Qy	179	ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp	198
Db	8080	GTGGACCCACAGATGGCGCGCTGCTGCTGGCGCTCGCGCGATGCC---GCCCTGTGG	8136
Qy	199	ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr	218
Db	8137	CGTAGCGAGCATCGCGCGCGCCTGGGCGAAGTGGCTGCTTCCCGCAGCGCGCTG	8196
Qy	219	TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp	238
Db	8197	GCGGTGCCACCGCACGCAATCATGTTGGCGCGCAGCAGCGGTG---GGGATCGCTTT	8253
Qy	239	ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly	258
Db	8254	GTCTGTTCGACGCGCAGTGGCGCACACCGCGCACCGCCACCGCGCATCTACGTGGGC	8313
Qy	259	ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla	278
Db	8314	GTGTCCACCGAACACACAGCCTGTATGTTCGGAAGATGCCGCGCGCAGTGGGACCG	8373
Qy	279	ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis	298
Db	8374	GTGGCGGCGCAGCGCGCGCTGCGCGCGCAGCCACATGCGCGCGCGCAGCAGTGGCAC	8433
Qy	299	ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal	318
Db	8434	---TGTACTCTAGCTACCGCACCGCGCGCGCGCAGCTGTATGCGCGCGGAGCCTTG	8490
Qy	319	TrpLysPheSerValThrSerGlyThrTrpArgIleSerProValProSerThrAsp	338
Db	8491	GTGAATTCACCGCGCACAGGGCGCTGGCGCGAGATAGCCCGGATTCGCGAG---CCA	8547
Qy	339	ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr	358
Db	8548	GCCAGTGGCGGATTCGGCTGGGCGCGGTGGCGGTGATCCGCAACACCGCAGGTG	8607
Qy	359	IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp	378
Db	8608	CTGTTGGCGCAGCACCTTCGCGCGCGCGCAGCGCGCGCAGCAGTGTATCGACGCTGAT	8667
Qy	379	GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg	398

Db 8668 GGTGCAAGCACTGGCGCGCTTG-----TTGCGCGAT 8700
 Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProPro 418
 Db 8701 GCGGTGTTCGACCAAGCGCGCGCTGG---ACCGCACATGCCACCGCAC----- 8751
 Qy 419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 438
 Db 8752 -----TGGATGGG---GCGCTGGGATCGATCCGTTGACCGC 8787
 Qy 439 AspArgMetLeuTrpGlyThrGlyAlaThrLeuTrpAlaThrAsnAspLeuThrLysTrp 458
 Db 8788 AACACGCGCTGTTCGTACCGCTACCGCATCTGGGCTCGCGCAATCTG----- 8838
 Qy 459 AspSerGlyGlnIleHisIleAlaPro-----MetVal 470
 Db 8839 -----CAGGATTCGCCCGCAGCGCGCGCTGCAGTGGTGTTCAGGAC 8886
 Qy 471 LysGlyLeuGluThrAlaValAsnAspIleSerProSerGlyAlaProLeu 490
 Db 8887 CGTGGCTTGGAGAAACCGTCCGCTGACCTGCTCAGCCCGATCGATCGCGCGCGCATCTG 8946
 Qy 491 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 510
 Db 8947 CTCAGCGCATCTCGCGCATATCATGCTTCGGCATGACGATTGGACCGCGC---CAG 9003
 Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTrpAlaGluLeuAsn 530
 Db 9004 TTGCAGTACCGCGCCCGCGCTGACCAATGCGAAGCATCGATCGCGCGCGCGCGCG 9063
 Qy 531 ProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550
 Db 9064 CGCAGTGGTGTGGCGAGCGGTACCGTG---CGCGACCGCGCGCAACAAATTCGCG 9120
 Qy 551 ValAlaPheSerThrAspGlyGlyValAsnTrpPheGln---GlySerGluProGlyGly 569
 Db 9121 GCGGTGTATTCGCGTATGGCGCGAAGCAATGACGGCGTTTGCAGTGGAGCGCGCGCG 9180
 Qy 570 ValThrThrGlyGlyValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589
 Db 9181 GGGCAGGCGCGCGCAGCATTCATGCTGGTGGATGGCGCGCAGTGGTGGTGGCGCGCG 9240
 Qy 590 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 609
 Db 9241 GAGCGCGCGCGCAAT-----TGGCGCACGTCGCACTTCGCGCGCGCAGTGGCGCGGTG 9294
 Qy 610 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629
 Db 9295 GACGCGCTCGCGACACCGCGGTGGTGGTGGCGATGCGCGACGCGCGCTGGTAT 9354
 Qy 630 Ala-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGln 647
 Db 9355 GCGGTGGATGTCGCCAGCGCGCGAGCTGTACGAGACGACCGATCCCGCGCGCATTCGCT 9414
 Qy 648 ProValAlaAlaGlyLeu----- 653
 Db 9415 -----GCGACCGGTGTGAGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCA 9468
 Qy 654 -----ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGly 671
 Db 9469 CTGCGTCCGCGACCGCGCGCGCGCGCGGTGTATCTGGCGACGCGCGCGCGCAAG----- 9522
 Qy 672 AspLeuTrpLeuAlaAlaSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrp 691
 Db 9523 -----GGTGTGATCGCTGGCAGGACGCT-----ACGCTG 9552
 Qy 692 SerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGly 711
 Db 9553 CAGGTGCTCTCACAGCGCGCGAGCAAGCGCTCGCTCGCATCGCGCAAGCGCTGCGCGCGC 9612
 Qy 712 SerSerTyrProAlaValPheValGlyThrIleGlyGlyValThrGlyAlaTyrArg 731
 Db 9613 GTGGCGCGCGCGCTGTATCTGCGCGCGCGCGCTGTCAGGCGCGTGGATGGCGGTTCGCG 9672

Qy 732 SerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTrpGlyAsn 751
 Db 9673 TCCGACGATGGCGGTGCAGTGGCAACCATCAACGATGACGCGCACCGTTCGGCGCG 9732
 Qy 752 TrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTrpIleGlyThrAsn 771
 Db 9733 ---CGGTACAGCGTACCGGTGATCCGCGCATTCGCGCGCGTGTGTACTTCGCCACCGCG 9789
 Qy 772 GlyArgGlyIleValTrpGlyAspIleGlyAlaProSerGlySerProSerProSer 791
 Db 9790 GCGCGCGCATTTCTACGCGCATCCGAGATGAGCAGTGCATGCCCTGCTCCCGA----- 9843
 Qy 792 ValSerProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer 810
 Db 9844 ---CGCCCTGCTGCTTACCGTGTACGGCGCGCGCGCGCTGGCGCGCGCGCGCGTGA 9900
 Qy 811 ---SerProSerProSerProSerProSerProSerProSerProSerProSerProSer 827
 Db 9901 CACCGCGCGACCTGGAGATCACCGCATGACCCCGT-CCACATTTCCGCGCGCTGCTTGC 9959
 Qy 828 -----ProSerProSerPro----- 832
 Db 9960 TGGGT 10019
 Qy 833 -----SerProSerArgSerProSerProSerProSerProSerProSerProSerPro 846
 Db 10020 CCGATGGCGCGGT 10079
 Qy 847 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 865
 Db 10080 GCGCTTCAA---TCAGGTGAGTTTCGATGGCAAGCAGGCGCACCGCCACAGACCGACG 10133

Search completed: October 5, 2005, 16:22:25
 Job time : 13428.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 06:17:10 ; Search time 1472.27 Seconds
(without alignments)
3847.935 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLTRSRRLVSL.....RASFGSYNPATPTADTVLQX 957

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame; p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool/US09917376/runat 04102005.164327 26889/app query.fasta.1.2524
-DB=N_Geneseq_16Dec04 -QWT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CGN 1.1.1867 @runat 04102005.164327 26889 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	2869	10 ADD22922	Add22922 Acidother
2	5134	100.0	2869	10 ABZ77632	Abz77632 Nucleotid
3	5130	99.9	2869	12 ADO52314	Ado52314 Acidother
4	2174	42.3	2823	8 ABV76941	Abv76941 Nucleotid
5	1631.5	31.8	2517	12 ADH51583	Adh51583 T reesei

6	1631.5	31.8	2710	12	ADH51580	Adh51580 Trichoder
7	1558.5	30.4	2849	10	ADD42060	Add42060 Trichoder
8	1167.5	22.7	3668	10	ADD24893	Add24893 DNA encod
9	1125.5	21.9	2646	12	ADH19117	Adh19117 Geotrichu
10	1124	21.9	2481	12	ADH19123	Adh19123 Geotrichu
11	1117	21.8	2367	12	ADH19119	Adh19119 Geotrichu
12	993.5	19.3	2755	13	ADR90295	Adr90295 Geotrichu
13	979	19.1	2268	13	ADR90297	Adr90297 Geotrichu
14	979	19.1	2274	13	ADR90301	Adr90301 Geotrichu
15	940	18.3	5698	10	ADD42054	Add42054 Trichoder
16	823	16.0	3687	8	ABZ77634	Abz77634 Nucleotid
17	823	16.0	3687	12	ADH36637	Adh36637 DNA encod
18	702	13.7	2289	8	ABZ77633	Abz77633 Nucleotid
19	702	13.7	2289	10	ADJ38292	Adj38292 A. cellul
20	429	8.4	3365	8	ABZ76162	Abz76162 A. cellu
21	409	8.0	11706	2	AAX55661	Aax55661 DNA seque
22	409	8.0	11707	6	AAD26525	Aad26525 Active ce
23	402	7.8	3364	10	ADF75896	Adf75896 Acidother
24	392.5	7.6	1103	3	AAF14988	Aaf14988 Trichoder
25	365.5	7.1	651	5	AAS84956	Aas84956 DNA encod
26	345	6.7	18596	4	AAF31109	Aaf31109 Thymidyla
27	345	6.7	18596	5	AAC91215	Aac91215 Human thy
28	345	6.7	18596	6	ABL67927	Ab167927 Ovary can
29	345	6.7	18596	6	ABL62854	Ab162854 Breast ca
30	345	6.7	18596	6	ABL63078	Ab163078 Breast ca
31	345	6.7	18596	6	ABK43334	Abk43334 Human Thy
32	345	6.7	18596	6	ABN95092	Abn95092 Gene #159
33	334.5	6.5	6415	2	AAX55662	Aax55662 DNA seque
34	334	6.5	6416	6	AAD26526	Aad26526 Active ce
35	330.5	6.4	7407	8	ACA42281	Acad42281 Prokaryot
36	330.5	6.4	7407	9	ACC59398	Acc59398 Microbial
37	323	6.3	2600	2	AAQ15178	Aaq15178 Portion o
38	319.5	6.2	8211	11	ABD15052	Abd15052 Peudomon
39	318	6.2	1080	6	ABI99537	Abi99537 Mouse isc
40	302	5.9	1050	4	AAD21685	Aad21685 Mutaciona
41	302	5.9	2849	4	AAD21684	Aad21684 Human ret
42	299.5	5.8	210528	11	ACN44040	Acn44040 Mouse gen
43	296.5	5.8	4767	3	AAA14667	Aaa14667 Nucleotid
44	293	5.7	4818	3	AAA14669	Aaa14669 Nucleotid
45	290	5.6	1232	4	AAD11124	Adad11124 Human sma

ALIGNMENTS

RESULT 1
ADD22922
ID ADD22922 standard; DNA; 2869 BP.
XX
AC ADD22922;
XX
DT 15-JAN-2004 (first entry)
XX
DE Acidothermus cellulolyticus avicelase AvIII DNA.
XX
KW AvIII; cellulose reduction; agricultural biomass; municipal solid waste;
KW glycoside hydrolase; avicelase; ds; gene.
XX
OS Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
FT CDS 1..2868
FT /tag= a
FT /product= "AvIII"
XX
PN US2003108988-A1.
XX
PD 12-JUN-2003.
XX
PF 18-OCT-2002; 2002US-00155400.
XX
PR 28-JUL-2001; 2001US-00917376.
XX
PA (DING/) DING S.

PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
XX (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel ME;
DR WPI; 2003-810853/76.
XX P-PSDB; ADD22921.

XX New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding Aviii and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX
PS Claim 17; SEQ ID NO 2; 29pp; English.
XX
CC The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable Aviii polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding Aviii. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents DNA encoding the Acidothermus cellulolyticus
CC avicelase Aviii.
XX
SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 1,46e-172 Length: 2869
Score: 5134.00 Matches: 956
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.98% Indels: 0
DB: 10 Gaps: 0

US-09-917-376-1 (1-957) x ADD22922 (1-2869)

QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu 20
DB 1 ATGGATCGTTCGGAGAACATCGCTGACTATGAGATCAGCAGCATTTGGTATCACTGCTC 60

QY 21 AlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThr 40
DB 61 GCCGCCACTGCGTCTGCGCGTGGCGCGCTCTGGGAGTCTGCCCATCGCGATAACG 120

QY 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60
DB 121 GCTTCTCTGCGCAGCGCGGACGACTCAGCGGTACACTGGAGCAACGTGGCGATCGGG 180

QY 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
DB 181 GCGCGCGCTTGTGCGACGGGATCGTCTCAATGAAGGTGCACCGGGAATCTGTACGCTG 240

QY 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyValTrpIleProLeu 100
DB 241 CGGACGACATCGGGGGATGTATCGATGGGATGCGCCAAACGGCGGTGGATCCCTTT 300

QY 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
DB 301 CTGGATTTGGGTGGATGGACAAATTTGGGGGTACACCGGGTCTGTACGCAATTCGGCGAC 360

QY 121 ProIleAsnThrAsnLysValTrpAlaAlaValAlaGlyMetTyrThrAsnSerTrpAspPro 140
DB 361 CCGATCAATACTAAACAGTATGGCGCGCTCGGAATGTACACCAACAGCTGGGACCCA 420

QY 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
DB 421 AACACGGAGCGATTTCTCGCTCGTCTGATCAGGGCGCAACGTGGCAATTAACGGCCCTG 480

QY 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
DB 481 CCGITCAAGCTTGGCGGCAACATGCCCGCGCTGGAAATGGCGGCGCGCTTCGGTGGAT 540

QY 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200
DB 541 CCAACAATGACAACTTCTGTATTTGGCGCGCCCGAGCGGCAAGGGCTCTGGAGAGC 600

QY 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220
DB 601 ACAGATTCCGGCGGACCTGGTCCAGATGACCACTTTCCGAGCGTAGCAGCTACATT 660

QY 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
DB 661 GCAATCCCACTCAGACACCGGCTATCAGACGATATCAAGGCGCTCTGTGGTGGCT 720

QY 241 PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla 260
DB 721 TTCGCAAGTCTTCTGTCATCGCTCGGCAAGCAGTAAGACCAATTTTGTGGCGGTGGCG 780

QY 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280
DB 781 GATCCCAATAATCCGCTCTTCTGGAGCAGAGACGGCGCGCGAGCGGTGCCG 840

QY 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300
DB 841 GGTGGCGCGACCGGCTTCATCCGCACAAAGGGCGTCTTTGACCCGCTCAACCACTGCTC 900

QY 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyValTrpLys 320
DB 901 TATATTGCCACAGCAATACGGGTGTCGTATGACGGGAGCTCCGCGCAGCTTGGAAA 960

QY 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340
DB 961 TTCTCGGTGACCTCCGGGACATGGACGCAATCAGCCCGGTACTTTCGACGGACACGGCC 1020

QY 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360
DB 1021 AACGACTACTTTGGTTACAGCGGCTCTCTATCGACCGCAGCAGCCGCAACGATAATG 1080

QY 361 ValAlaThrGlnIleSerTrpProAspThrIleIlePheArgSerThrAspGlyGly 380
DB 1081 GTGGCAACCCAGATATCGGTGGTGGCGGACACCAATATCTTTCGAGACCGGACGGCGT 1140

QY 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400
DB 1141 GCGAGCTGGACCGCGGATCTGGGATGGACGAGTTATCCCAATCGAAGCTTGCATATGTG 1200

QY 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProValPro 420
DB 1201 CTTGACATTTCCGCGGAGCTTGGCTGACCTTCGGCGTACAGCCGAATCTCCCGTACCC 1260

QY 421 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440
DB 1261 AGTCGGAAGCTCGGCTGGATGGATGAAGCGATGGCAATCGATCCGTTCAACTCTGATCGG 1320

QY 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460
DB 1321 ATGCTCTACGGAACAGCGCGGACGTTGTACGCAACAAATGATCTCAGGAAGTGGGACTCC 1380

QY 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 480
DB 1381 GCGCGCCAGATTCATATCGCGCGGATGGTCAAAGGATTCGAGGAGACGGCGGTAAACGAT 1440

QY 481 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500
DB 1441 CTCATCAGCCCGCGCTGCGCGCCCGCTCATCAGCGCTCTCCGAGACCTTCGCGCGGCTTC 1500

QY 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520
DB 1501 ACCCAGCGGAGCTTACTGCGGTGCCATCGCATCTTACGTCACCGGTGTTACGAGCC 1560

QY 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540
DB 1561 GGCACACGCGTGCATATCGGGAATGAATCCGTCGATCATCGTTCGCGCTGGAAGTTTC 1620

QY 541 AspProSerSerGlnProAenAspArgHisValAlaPheSerThrAspGlyGlyLeuAen 560
DB 1621 GATCATCGAGCCCAACCGAGCAGGACGTCGCGTTCGACAGAGCGCGGCAAGAAC 1680
QY 561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 580
DB 1681 TGGTTCGAAGCAGCAGCACTGGCGGGGTGACGAGCGCGCACCGTCGCGCATCGGCC 1740
QY 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValThrAlaVal 600
DB 1741 GACGGCTCTCGTTCGTCGCGCTCCCGCGCATCCCGGTGAGCCTGTGTGTGTCAGCAGTC 1800
QY 601 GlyPheGlyAenSerTrpAlaAlaSerGlnGlyValProAlaAenAlaGlnIleArgSer 620
DB 1801 GGATTGGCAACTCTCGGGCTGCTTCGCAAGGTGTTCGCCCAATGCCAGATCCGCTCA 1860
QY 621 AspArgValAenProLysThrPheTyrAlaLeuSerAenGlyThrPheTyrArgSerThr 640
DB 1861 GACCGGGTGAATCCAAAGACTTCTATGCCCTTCCATGGACCTTCTATCGAAGCAG 1920
QY 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660
DB 1921 GACGCGCGGTGACATTCACCGCGTCCGCGCGGTCTTCCGAGCAGCGGTGCGCTCGGT 1980
QY 661 ValMetPheHisAlaValProGlyLysGlnGlyAspLeuTrpLeuAlaAlaSerSerGly 680
DB 1981 GTCATGTTCCACGCGGTGCTGGAAAGAGAGCGATCTGTGCTCGCTGCAATCGAGCGGG 2040
QY 681 LeuTyrHisSerThrAenGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700
DB 2041 CTTTACCACCTCAACCAATGGCGGAGCAGTTGGTCTGCAATCACCGGGTATCTCCGGG 2100
QY 701 ValAenValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720
DB 2101 GTGAACGTGGGATTTGGTAAGTCTGCGCGCGGTGCTCATACCCAGCGTCTTTGTGTC 2160
QY 721 GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpVal 740
DB 2161 GGCACGATCGGAGGGGTTACGGGGGCGTACCGCTCCGACGACTGTGGGACGACCTGGGTA 2220
QY 741 LeuIleAenAspAspGlnHisGlnTyrGlyAenTrpGlyGlnAlaIleThrGlyAspHis 760
DB 2221 CTGATCAATGATGACGAGCACCACCAATACGGAATTTGGGACAAAGCAATCACCGGTGACCC 2280
QY 761 AlaAenLeuArgArgValTyrIleGlyThrAenGlyArgGlyIleValTyrGlyAspIle 780
DB 2281 GCGAATTTACGGGGGTGTAGATAGCAGCAAGCGCGGTGGAAATTTGTATACGGGACATT 2340
QY 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800
DB 2341 GGTGTGTGCGCGTTCGCGATCGCGCTCTCCGTGGTGAGTCCGTGCGCTTCGCGGACGCTG 2400
QY 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
DB 2401 AGCCCGAGCCGAGCCGAGCAGCTCCCATCGCGTCCCGTCCCGCGGAGCTCGAGTCCA 2460
QY 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
DB 2461 TCTCTGTGCGCGTCTCCGTCGCGGTCCCATCCATCCGAGTCCGTCTCGGTTCGTCACCA 2520
QY 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
DB 2521 TCGCGGTTCGCGAGCCGCTTCTCGTCACGAGCCGCTCTCGTACCGCTTCGTCGCGCG 2580
QY 861 SerProThrProSerSerSerProValSerGlyValIleValGlnTyrIleAenAen 880
DB 2581 AGCCCAACCGCGTCTGTCGCGCGGTGTCGGGTGGGGTGAAGGTGACAGTATAGAATAAT 2640
QY 881 AspSerAlaProGlyAspAenGlnIleLysProGlyLeuGlnValValAenThrGlySer 900
DB 2641 GATTCCGCGCGCGGTGATTAATCAAGATCAAGCCGGGTTTCAGGTGTGTGAATACCGGGTGC 2700
901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920

DB 2701 TCGTCGGTGGATTTCGTCGACGCTGACGGTGGCGTACTGCTTCCACCGGATCGTGGCTCG 2760
QY 921 SerThrLeuValTyrAenCysAspTrpAlaIleGlyCysGlyAenIleArgAlaSer 940
DB 2761 TCGACACTGGGTGTACACTGTGACTGGCGCGGATCGGGTGTGGGAAATTCGCGGCTCG 2820
QY 941 PheGlySerValAenProAlaThrProThrAlaAspThrTyrLeuGln 956
DB 2821 TTCGGCTCGGTGAACCCGCGGACGCGGCGGACACCTACTCTGCAG 2868
RESULT 2
ID ABZ77632 standard; DNA; 2869 BP.
XX AC ABZ77632;
XX 03-JUN-2003 (first entry)
XX Nucleotide sequence of the avicelase AviIII.
XX Avicelase; AviIII; glycoside hydrolase; enzyme; cellulase; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
KW cellulose; gene; ss.
XX Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
CDS 1..2869
FT /*tag= a
FT /partial
FT /product= "AviIII"
FT /transl_except= (pos:2869,aa:Xaa)
FT /note= "Xaa is an unspecified residue"
XX
XX WO2003012090-A2.
XX
XX 13-FEB-2003.
XX
XX 28-JUL-2001; 2001WO-US023818.
XX
XX 28-JUL-2001; 2001WO-US023818.
XX (WIDE) MIDWEST RES INST.
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI: 2003-248177/24.
XX P-PSDB; ABP73015.
XX
XX New thermostable AviIII peptide from Acidothermus cellulolyticus, useful
PT for degradation of cellulose or in generating anti-AviIII antibodies for
PT purifying recombinant AviIII polypeptides from genetically engineered
PT host cells.
XX
XX Claim 3; Page 24; 44pp; English.
XX
XX The present sequence encodes a thermostable avicelase polypeptide,
CC designated AviIII. AviIII is a member of the glycoside hydrolase family
CC of enzymes, and is a cellulase. AviIII is useful in the conversion of
CC biomass to biofuels and biofuel additives. It may be useful in the
CC production of detergents, pulp and paper processing, food and feed
CC processing and in textile processes. The thermostable AviIII peptide is
CC useful in the degradation of cellulose, and in generating specific anti-
CC AviIII antibodies that are useful in purifying recombinant AviIII
CC polypeptides from genetically engineered host cells, in detecting AviIII
CC polypeptide expression, as well as a reagent tool for characterizing the
CC molecular actions of the polypeptide. The AviIII polynucleotide is useful
XX as a source of probes or primers in various diagnostic assays
XX
XX Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;
XX
XX Alignment Scores:

Pred. No.:	1.46e-172	Length:	2869
Score:	5134.00	Matches:	956
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.98%	Indels:	0
DB:	10	Gaps:	0
US-09-917-376-1 (1-957) x ABZ77632 (1-2869)			
QY	1	MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu	20
DB	1	ATGGATCGTTCGGAGAAATCCGTCGTGACTATGAGATCAGCACGATGGTATCACTGCTC	60
QY	21	AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr	40
DB	61	GCGGCACCTGGCTCGTTGCCGTGGCCGCGCTCTGGAGATTCTGCCCATCGCGATAAG	120
QY	41	AlaSerProAlaHisAlaAlaThrGlnProIleThrTrpSerAsnValAlaIleGly	60
DB	121	GCTTCTCTGCGCACGCGGAGCACTCAGCGGTACACTGGAGCAACGTTGGCGATCGG	180
QY	61	GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuThrVal	80
DB	181	GCGCGCGCTTGTGCGCGGATCGTCTTCAATGAAGGTGCACCGGGAATTCGTACGTG	240
QY	81	ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyValGlyTrpIleProLeu	100
DB	241	CGGACGGACATCGGGGGGATGATCGATGGGATGCCCAACGGCGGTGGATCCCTCTT	300
QY	101	LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp	120
DB	301	CTGGATTGGGTGGATGAGACAAATTTGGGGGTACACGGCGTCTGTCAGCATTTGCGG	360
QY	121	ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro	140
DB	361	CCGATCAATACTAAACAAGGTATGGCGCGCTCGGAATGTACACCAACAGCTGGGACCCA	420
QY	141	AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu	160
DB	421	AACGACGGAGGAGATTCCTCGCTCGTCTGATCAGGGCGCAACGTGGCAATAACGCCCTG	480
QY	161	ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp	180
DB	481	CGTTCAAGCTTGGCGGCAACATGCCCGGGGTGGAATGGGCGAGCGGCTTGGGTGGAT	540
QY	181	ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer	200
DB	541	CCAAACAATGACAACTTCTGATTTTCGGCGCCCGGAGCGGCAAGGGGCTCTGGAGAAGC	600
QY	201	ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle	220
DB	601	ACAGATTCGGCGGACCTGGTCCAGATGACGAATCTTTCGGAGCGTAGGCACGTACACTT	660
QY	221	AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla	240
DB	661	GCAATCCCACTGACACGCCGCTATCAGACGATATTCAGGCGTGTCTGGGTGCT	720
QY	241	PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla	260
DB	721	TTCCGACAAGTCTTCGTATCTCGTCGGCAAGCGAGTAAGACCAATTTTGTGGCGTGGCG	780
QY	261	AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro	280
DB	781	GATCCCAATAATCCGGTCTCTTCGGAGCAGACGCGCGGCGACGTTGGCAGCGGTGGCG	840
QY	281	GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu	300
DB	841	GSTGGCGCACCGGCTTCAATCCCGCACAGGGCGCTTTTGACCCGCTCAACCACGTGCTC	900
QY	301	TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys	320
DB	901	TATATTGCCACCAGCAATACGGGTGGTCCGTATGACGGGAGCTCCGGCGACGCTCGGAAA	960
QY	321	PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla	340
DB	961	TTTCTCGGTGACCTCCGGGACATGGACGGAAATCAGCCCGGTACCTTCGACGGACACGGCC	1020
QY	341	AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet	360
DB	1021	AACGACTACTTTGGTTACAGCGGCTCTACTATCGACCGCGACGCCCAACAGATAATG	1080
QY	361	ValAlaThrGlnIleSerTrpProAspThrIleIlePheArgSerThrAspGlyGly	380
DB	1081	GTGGCAACCCAGATATCTGGTGGCGGACACATAATCTTTCGGAGCACCGACGGCGT	1140
QY	381	AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal	400
DB	1141	GCGACGTGGACGCGGATCTGGGATTTGGACAGATTATCCCAATCGAAGCTTGCATATGTG	1200
QY	401	LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro	420
DB	1201	CTTGACATTTCCGGCGAGCTTTGGCTGACCTTCGGCGTACAGCCGAATCTCCCGTACCC	1260
QY	421	SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg	440
DB	1261	AGTCGAAGCTCGCTGGATGGATGAAGCATGGCAATCGATCCGTTCAACTCTGATCGG	1320
QY	441	MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer	460
DB	1321	ATGCTCTACGGAACAGGCGCGACGTTGTACGCAACAAATGATCTCACGAAGTGGGACTCC	1380
QY	461	GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp	480
DB	1381	GGCGGCGAGATTCATATCGCGCGGATGGTCAAGGATTTGGAGGAGACGGCGTAAACGAT	1440
QY	481	LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe	500
DB	1441	CTCATCAGCCCGCGCTCTGGCGCCCGCTCATCAGCGCTCTCGAGACCTTCGGCGGCTTC	1500
QY	501	ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr	520
DB	1501	ACCCACGCGAGCTTACTGCGCGTGCCATCGACGATCTTCACGTCACCGGTGTTCACGACC	1560
QY	521	GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe	540
DB	1561	GGCACCGCGTCACTATCGGAAATTTGAATCCGTCGTCATCATCTCGCGCTGGAAGTTTC	1620
QY	541	AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn	560
DB	1621	GATCCATCGAGCAACCCGAACACAGGACGTCGCGTCTTCGACACAGCGCGGCAAGAC	1680
QY	561	TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla	580
DB	1681	TGGTTCCAAGGACGACGACCTGGCGGGGTGACGACGGGGGACCGTCGCGCATCGGCC	1740
QY	581	AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal	600
DB	1741	GACGGCTCTCGTTTCGTTCGGCTCCCGCGCATCCCGGTCAAGCTGTGTGTACGCAATC	1800
QY	601	GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer	620
DB	1801	GGATTTGGCAACTCTCTGGGCTGTTCGCAAGGTGTTCCCGCAATCCCAAGATCCCGTCA	1860
QY	621	AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr	640
DB	1861	GACCGGTTGAATCCAAAGACTTTCTATGCTTATCCATATGGAACCTTCTATCGAAGCAG	1920
QY	641	AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly	660
DB	1921	GACGGCGGCTGACATTCACACCGGTCCGCGCGGCTCTTCGAGCAGCGGTGCGCTCGGT	1980
QY	661	ValMetPheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaAlaSerSerGly	680
DB	1981	GTCAATTCACGCGGTGCTTGAAAGAAGGCGATCTGTGGCTCGCTGATCGAGCGGG	2040
QY	681	LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla	700


```
Db 2041 CTTTACCACTCAACCAATGGCGGACGAGTGGTCTGCAATCACGGCGTATCTCCCG 2100
Qy 701 ValAenValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720
Db 2101 GTGAAGCTGGGATTTGGTAAAGTCTGCGCCCGGCTCGTCAATACCCAGCGCTTTGTGCTC 2160
Qy 721 GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrVal 740
Db 2161 GGCACGATCGGAGCGGTTACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTA 2220
Qy 741 LeuIleAenAspAspGlnHisGlnTyrGlyAenTyrGlyGlnAlaIleThrGlyAspHis 760
Db 2221 CTGATCAATGATGACGACCAACATACGGAATATGGGACACAGCAATACCGGTGACCA 2280
Qy 761 AlaAenLeuArgArgValTyrIleGlyThrAenGlyArgGlyIleValTyrGlyAspIle 780
Db 2281 GCGAATTTACGGCGGTTGATACATAGCAGCAACGCGCTGGGAATTTGATACGGGACATT 2340
Qy 781 GlyGlyAlaProSerGlySerProSerProSerProSerProSerProSerProSerPro 800
Db 2341 GGTGTGCGCGCTCCGGATCGCGTCTCGCTCGGTAGTCCGTCCGCTTCGCCGACCTG 2400
Qy 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
Db 2401 AGCCCCGAGCCGAGCCGAGCAGCTCGCCATCGCCGTCCGCTCGCGGAGCTCGAGTCA 2460
Qy 821 SerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 840
Db 2461 TCTGTGTGCGCTCTCGCTCGCGCTCACCATCGCCGAGTCCGCTCTCGGTCCGTCACCA 2520
Qy 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
Db 2521 TCGCGCTCGCGAGCCGCTCTTCTCGTCACCGAGCCGCTCTCGTCACCGCTTCGTCGCG 2580
Qy 861 SerProThrProSerProSerProValSerGlyGlyValIleValGlnTyrIleAen 880
Db 2581 AGCCCAACGCGCTCGTCTCGCGGTGCGGTGGGTGAGGTGAGTATAGGAATAAT 2640
Qy 881 AspSerAlaProGlyAspAenGlnIleLysProGlyLeuGlnValValAenThrGlySer 900
Db 2641 GATTTCGGCGCCGGGTGATTAATCAGATCAAGCCGGTTTCAGGTGGTGAATACCGGTG 2700
Qy 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920
Db 2701 TCGTGTGTGATTTGTGCGAGGTGACGCTGCGTACTGTTTACCGCGGATGCTGCTCG 2760
Qy 921 SerThrLeuValTyrAenCysAspTrpAlaAlaIleGlyCysGlyAenIleArgAlaSer 940
Db 2761 TCACACTGCTGTACAACTGTGACTGGCGGCGATCGGGTGGGAATATCCGCGCTCG 2820
Qy 941 PheGlySerValAenProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 2821 TTGCGTGTGTTGAACCCGCGAGCCGACCGCGGACACCTACTCTGCGAG 2868

RESULT 3
ID ADO52314 standard; DNA; 2869 BP.
XX AC ADO52314;
XX DT 15-JUL-2004 (first entry)
XX DE Acidothermus cellulolyticus avicelase III (AviIII) DNA.
XX KW Thermostable cellulase; avicelase III; AviIII; cellulose reduction;
KW KM biomass degradation; ethanol formation; industrial chemical;
KW KX fabric treatment; gene; ds.
XX OS Acidothermus cellulolyticus.
XX FH Key Location/Qualifiers
FT CDS 1..2868
```

```
/*tag= b
/product= "Acidothermus cellulolyticus avicelase III"
/notes= "CDS contains translational exceptions"
/partial
/notes= "No stop codon"
1..306
/*tag= a
307..2865
/*tag= c
/product= "Acidothermus cellulolyticus mature avicelase
III"
/*tag= d
/notes= "A string of unknown nucleic acid units"
2869
US2004038334-A1.
26-FEB-2004.
28-JUL-2001; 2001US-00917376.
28-JUL-2001; 2001US-00917376.
(DING/) DING S.
(ADNE/) ADNEY W S.
(VINZ/) VINZANT T B.
(HIMM/) HIMMEL M E.
Ding S, Adney WS, Vinzant TB, Himmel ME;
WPI: 2004-203224/19.
P-PSDB; ADO52313.
Novel thermostable AviIII polypeptide of glycoside hydrolase family and
isolated from Acidothermus cellulolyticus, useful for degrading treated
biomass into simpler forms of carbohydrate.
Claim 17; SEQ ID NO 2; 19pp; English.
The invention relates to a thermostable cellulase enzyme, avicelase III
(AviIII) and its nucleic acid sequence. AviIII is useful for reducing
cellulose in a starting material. A thermostable AviIII peptide is useful
for degrading treated biomass into simpler forms of carbohydrate, which
is used in the formation of ethanol or other industrial chemicals. It is
also useful for treating fabrics to remove cellulose-containing stains.
The present sequence is Acidothermus cellulolyticus AviIII DNA.
Sequence 2869 BP; 546 A; 857 C; 899 G; 565 T; 0 U; 2 Other;
Alignment Scores:
Pred. No.: 2,02e-172 Length: 2869
Score: 5130.00 Matches: 955
Percent Similarity: 99.90% Conservative: 0
Best Local Similarity: 99.90% Mismatches: 1
Query Match: 99.90% Indels: 0
DB: 12 Gaps: 0
US-09-917-376-1 (1-957) x ADO52314 (1-2869)
Qy 1 MetAspArgSerGluAenIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu 20
Db 1 ATGGATCGTTCCGGAGAACATCCGCTGACTATGAGTACACGACGATGGTATCACTGCTC 60
Qy 21 AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr 40
Db 61 GCGCCCACTGGTCTCGCGGTGGCGCGCTCTGGGAGTCTTCCCATCGGATAACG 120
Qy 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60
Db 121 GCTTCTCTCGCAGCGCGGCGGACCTCAGCCGTACACCTCGGAGCAACGTCGCGATCGGG 180
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
```


Db 181 GCGCGCGCTTTGTTCGACGGGATCGTCTTCAATGAAGGTGCACCGGGAATTTCTGTACGTG 240
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAenGlyValTrpIleProLeu 100
Db 241 CGGACGGACATCGGGGGGATGTATCGATGGGATGCCGCCAACGGCGGTGGATCCCTCTT 300
Qy 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
Db 301 CTGGATTGGTGGATGGNAACAATTTGGGGGTACACACGGCGTCTGCAGCATTTGGCGCAGAC 360
Qy 121 ProfileAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140
Db 361 CCGATCAATACTAAACAGGTATGGCGCGCTCGGAATGTACACAAACAGCTGGGACCCA 420
Qy 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
Db 421 AACGACGGAGCGATTCTCCGCTCGTCTGATCAGGGCGCAAGGTGGCAATTAACGGCCCTG 480
Qy 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
Db 481 CCGTTCAAGCTTGGCGGCAACATGCCCGGGCGTGGATGGGCGAGCGCTTGGCGTGGAT 540
Qy 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200
Db 541 CCAACAATGACAAATTTCTGATTTTCGGCGCCCCGAGCGGCAAAAGGCTCTGGAGAAAGC 600
Qy 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220
Db 601 ACAGATTCCGGCGCACCTTGGTCCAGATGACGAATTTCCGGACGTAGGCACGTACATT 660
Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
Db 661 GCAATCCCATCTGACACACACCGGCTATCAGACCGATATTCAGGCGTCTGTGGGTGCT 720
Qy 241 PheAspLysSerSerSerLeuGlnAlaSerLysThrIlePheValGlyValAla 260
Db 721 TTCGACAAAGTCTTCTGTCATCGCTCGGCAAGCGAGTAAAGACCATTTTTGTGGCGTGGCG 780
Qy 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280
Db 781 GATCCCAATAATCCGGTCTTCTGGAGCAGACGCGCGCGACGTGGCGCGGTGGCG 840
Qy 281 GlyAlaProThrGlyPheIleProHissGlyValPheAspProValAsnHissValLeu 300
Db 841 GGTGGCGCACCGGCTTATCCCGCACAGGGCGTCTTTGACCCCGTCAACACACGTGCTC 900
Qy 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 320
Db 901 TATATTGCCACCAGCAATACGGTGTGTCGTATGACGGGAGCTCCGGCGACGTCTTGAAA 960
Qy 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340
Db 961 TTCTCGGTGACCTCCGGGACATGGACCGGAATCAGCCCGGTACCTTCGACGGACACGGCC 1020
Qy 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHissProAsnThrIleMet 360
Db 1021 AACGACTACTTTGGTTACAGCGGCTCACTATCGACCGCGCACCCGAAACACGATAATG 1080
Qy 361 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 380
Db 1081 GTGGCAACCCAGATATCGTGGTGGCGGACCACTAATCTTTTCGGAGCACCCACGGCGGT 1140
Qy 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400
Db 1141 GCGACGTGGACCGCATCTGGGATTTGGACAGTATTCACCAATCGAAGCTTGGCATATGTG 1200
Qy 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420
Db 1201 CTTGACATTTTCGGCGAGCCTTTGGTGACCTTTCGGCGGTACAGCCGAATCTCTCCCGTACCC 1260
Qy 421 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440
Db 1261 AGTCCGAAGCTCGGTGGATGGATGAGCGGATGGCAATCGATCCGTTCAACTCTGATCGG 1320

Qy 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460
Db 1321 ATGCTCTACGGACACAGGCGCGAGCTGTGTACGCAACAATGATCTCACGAAGTGGACTCC 1380
Qy 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 480
Db 1381 GCGCGCCAGATTCAATATCGCGCGATGGTCAAAAGGATTGGAGGAGACGGCGTAAACGAT 1440
Qy 481 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500
Db 1441 CTCATCAGCCCGCTCTGGCGCCCGCTCATCAGCGCTCTCGAGACCTTCGGCGGCTTC 1500
Qy 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520
Db 1501 ACCACGCGCAGCTTACTGCGCGTGCCATCGACGATCTTCACTCAGCGGTGTTCCAGCAC 1560
Qy 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540
Db 1561 GGCACACGCGTCACTATCGGGAATTTGAATCCGTCGATCATCGTTCGCGCTGGAAGTTTC 1620
Qy 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560
Db 1621 GATCCATCGAGCAACCCGAACGACAGGCACGTCGCTTCTCGACACGCGCGCAAGAAC 1680
Qy 561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 580
Db 1681 TGGTTCCAAGGACGACCACTGGCGGGGTGACACGCGGGGACACCGTCGCGCATCGGCC 1740
Qy 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 600
Db 1741 GACGCTCTCTGTTTGTCTGGGCTCCCGCGCATCCCGGTCAAGCTGTGGTGTACGCAATC 1800
Qy 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620
Db 1801 GGATTTGGCAACTCTCTGGGCTGCTTCCCAAGGTGTTCCCGCCAATCCCAAGATCCCGTCA 1860
Qy 621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640
Db 1861 GACCGGTGTAATCCAAAGACTTTTCTATGCCCTATCCATGGAACCTTCTATCCAGGACG 1920
Qy 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660
Db 1921 GACGCGCGGTGACATTCACACGGTCCGCGCGCTTCCGAGCAGCGGTGCGTGGT 1980
Qy 661 ValMetPheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaAlaSerSerGly 680
Db 1981 GTCATGTTCCACGCGTGCCTGGAAAGAAGGCGATCTGTGGCTCGTGCATCGAGCGGG 2040
Qy 681 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700
Db 2041 CTTTACCACCTCAACCAATGGCGCAGAGTTGCTGTGCATACCCGCGGTATCTCTCCGCG 2100
Qy 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720
Db 2101 GTGACGTGGGATTTGGTAAGTCTGCGCCCGGCTCGTCATACCACGCGCTTTTGTGCTC 2160
Qy 721 GlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpVal 740
Db 2161 GGCACGATCGGAGCGCTTACGGGGGGGTACCGCTCCGACGACTGTGGGACGACCTGGGTA 2220
Qy 741 LeuIleAsnAspAspGlnHissGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760
Db 2221 CTGATCAATGTATGACAGCACCAATACGGAAATTTGGGGAACAACAATCACCGGTGACCA 2280
Qy 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 2281 GCGAATTTACGGCGGGGTGTACATAGGACGAAACGGCGGTGGAAATTTGTATACGGGGACAT 2340
Qy 781 GlyGlyAlaProSerSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800
Db 2341 GGTGGTGGCGCGTCCGGATCGCGCTCTCCGTCCGTGAGTCCGTCCGCTTCCCGAGCGCTG 2400

QY 249 GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp 268
DB 715 GGTGAGACACAAACCAACATCTACGTAGGTGTAGCAGACCAACGAAACAAAGTTTATCGA 774
QY 269 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro 288
DB 775 TCTACCGATGGGGGCAACATGTCGCAACGCGTCCCTGGACACACCCACCGGATTCCTGGCG 834
QY 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308
DB 835 CAAAGGGAGTCTTGACCATAAAGGTGACCAACTCTACATCGCTACCTCAGATACTGCT 894
QY 309 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp 328
DB 895 GCCCGTATGACGGGTGAGGGGATGTGGCCCTAGATATTTCCAGTGGCCAGTGG 954
QY 329 ThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGly 348
DB 955 ACTCGAATCAGCCCAATTCCTCTACC--TCATCGAATTCAGCGTTCGGGTACTCCGT 1011
QY 349 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp 368
DB 1012 TTAGCCATTGACCGCAAAACCTGACACCATCATGTTGTTTCCCAAGTGTCTTGTTGG 1071
QY 369 ProAspThrIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAsp 388
DB 1072 CTTGACATGTAGTATATCGCTCCACCGACCGGGGAAACCGTGTGCTTCTCTGCGAA 1131
QY 389 TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp 408
DB 1132 CTGACGGTTTCACACCGCGGTACAAAGCAGTACACACCGATTACTCAGGGGTCCTCTGG 1191
QY 409 LeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAsp 428
DB 1192 CTTGACTTTGGCAATACAGCCAAAGAGCCTGAAGCTAACCCCAACTCGGGTGGATGACC 1251
QY 429 GluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThr 448
DB 1252 CAGTCCTTCGAGATTGACCGCATAACTCCGACCGCTTCTTCTATGGCAGCGCGCGGCG 1311
QY 449 LeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro 468
DB 1312 ATCTATGGCGGACCAACCTCACAACCTGGGATAGGGCAAAAAGTTGATATCACTGTC 1371
QY 469 MetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAla 488
DB 1372 AAGGCCCAAGGATCGAAGAACCCGAGCCCAAGACCTTGGCGCTCCCTCGGAAACATT 1431
QY 489 ProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaVal 508
DB 1432 GCCCTATACTCAGCGCTGCGACATCGTGGCTTCACCCCAAGATATCTCTCAGGTG 1491
QY 509 ProSerThr--IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla 527
DB 1492 CTTAACAAGTACTACTACAGAACCCACCATGACACCGTGACCTCTCATGACTTCGCC 1551
QY 528 GluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsn 547
DB 1552 GAAAGCAAAACCGCAACTGCTGCTGCTGCGAGCAAA--TCTATTCTGGGGAACCC 1605
QY 548 AspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluPro 567
DB 1606 ACCTCATGGGTTGGGGTTTCAACTGATGCTGGCGAAACGTGGAAACCTGGCGCCACGCCA 1665
QY 568 GlyGlyValThrThrGlyGlyValAlaAlaSerAlaAspGlySerArgPheValTrp 587
DB 1666 TCAGGGGTTAAAGGCCCGCCCTCCATCACTGTGTCAGCTTAATGCTTCATCCATTGTGG 1725
QY 588 AlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAla 607
DB 1726 GCCCG-----GAAGGTGCCCTCTCGACGCTTCCACCAACAGTGGCTCTCTCGTGCTCA 1779

QY 608 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 627
DB 1780 ACGGTGACGGGCTTCTCACAATGCCAGGTCCGATCAGACCGAGTCAACGCTAATACT 1839
QY 628 PheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGln 647
DB 1840 CTTTACGGCTCGTTTGACGGAAAGTTCTATCACTCCACAAATGGTGGGCTTCTCTCACG 1899
QY 648 ProValAla--AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666
DB 1900 GCGTCAGCTTTCACGGGTTTCCACCACTGGA-----AACGTTCTGTTCCGTCGCGTG 1953
QY 667 ProGlyLysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeu 681
DB 1954 CCGGCGCTCAAGGTCACTCTGCGTTCGTGGAGTATCGGAAGACCTATGGGATG 2013
QY 682 TyrHisSerThrAsnGlyGlySerTrpSerAlaIleThrGlyValSerSerAlaVal 701
DB 2014 TGGCGGTCTCATGTATGGCGGGAAGAACTGGAACCAAGGTATCTCGAGTTTCAGGAAGCGAT 2073
QY 702 AsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGly 721
DB 2074 GCGGTGTTTTCGGTAAAGCAACCACTAGTAGCGGATACCCAGTGATTTTCAATCCGCC 2133
QY 722 ThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeu 741
DB 2134 AAGATTGATGAGTTTCGAGGAATTTTCGCTCTGACGATGAAGGCAAGACCTTGGAAACGC 2193
QY 742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761
DB 2194 ATCAACGACCAACCAACAGTGGTATGGACTGGCGCATCAATCACCGGTGACCCAGAT 2253
QY 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGly 781
DB 2254 GTCGAAGGCGCGCTACATTGGTACCAACCGTTCGCGCATATCTGTTGAT----- 2307
QY 782 GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer 801
DB 2307 ----- 2307
QY 802 ProSerProSerProSerSerProSerProSerProSerProSerProSerProSer 821
DB 2308 -----TCTCTCACTCCCGCCCCCGGTCAAGATAGCGCAGCATCG 2352
QY 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
DB 2353 GGAATATACAGGGGCCACACCGCGCTGAAACCCCTGCAGCGATCACAACACTGCGCGG 2412
QY 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSer 860
DB 2413 TCCACTGATACGGCGAAACTTCAGAGCCGGGAACAGGAAGTGGATCTCGAACCCGACCT 2472
QY 861 SerProThrProSerSerSerProValSerGlyValLysValGlnTyrLysAsnAsn 880
DB 2473 GACTCGGGGAACACCTCTCTGGGAATACCACTTCATGCAAGTTCGCTATTCACGACA 2532
QY 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
DB 2533 GACTGGGCGAGCGGC-----TTCACTGGGCGAGCGACCATCACTAACACGCTCCGCC 2583
QY 901 SerSerValAspLeuSerThrValThrValArgTyr----- 912
DB 2584 AGCCCAATTAACGGCTGGACGTTGCGGTTCATTACCTTCGGGGGAGACCATCACTTCC 2643
QY 913 ---TrpPheThrArgAsp-----GlyGlySerSerThrLeuValTyrAsnCysAspTrp 929
DB 2644 GTGTGGTCAGCAACTCAGACTCTGTGTCCTCAATGTGTGCTCAAGAACTCTGTTGG 2703
QY 930 AlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 949
DB 2704 -----AATCCACGATACCT 2718

RESULT 5

QY 379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398
 Db 1039 TCTGGGCAACATGAGCCCGATCTGGCGTGGCGAGCTATCCGACTGAGCACTATTAC 1098
 QY 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGly---ValGlnProAsnPro 417
 Db 1099 TACAGCATCTCACTCCCAAGACCGTGGATCAAGAACAACTTTATCGATGTGACGAGC 1158
 QY 418 ProValProSer-----ProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433
 Db 1159 GAGTCACCGTCCGATGCTCATCAAGCGCTCGCGTGGATGATCTCTCGAGATT 1218
 QY 434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453
 Db 1219 GACCAACCGGACCAACCATCTGGCTCTACGGCACCGGAATGACANTCTTTGGCGGCCAC 1278
 QY 454 AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValIysGlyLeu 473
 Db 1279 GATCTCACTGGGACACCGGCCACATGTGTCAATCCAATCACTGGCAGCGGCATC 1338
 QY 474 GluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAla 493
 Db 1339 GAGGAATCTCTCGTCAGGACCTGGCTCTGCACCGCGGAAGCGAGCTATTGGCCGCA 1398
 QY 494 LeuGlyAspLeuGlyGlyPheThrHisIle-----AspValThrAlaValProSer 510
 Db 1399 GTCCGAGACGACACCGGCTTCACTTTGCCAGCAGAAACGACCTCGGACATCGCCGAG 1458
 QY 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
 Db 1459 ACGTCTCGGCACACCCACATGGGCCACCTCGAGAGCGTTCGACTACCGCGGAACTCG 1518
 QY 531 ProSerIleValArgAlaGlySerPheAspProPheSerSerGlnProAsnAspArgHis 550
 Db 1519 GTCAGAGCGTGTCTCGCGTGGCAACACCGCGCGCAACAG-----1563
 QY 551 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal 570
 Db 1564 GTGGCCATCTCTGTCGCGCGCGCGCGAGCATCGACTACCGCGCGCAGACAGTCC 1623
 QY 571 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 590
 Db 1624 ATGAACGGCGGACGCGTGGCTATTTCGCCGACGCGCAGACATCTCTGTGTCGACCGCC 1683
 QY 591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
 Db 1684 TCGTCCGCG-----GTGCAGCGCTCGCAGTTCAGGGCAGCTTTGCTCCGCTCTCG 1734
 QY 611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 630
 Db 1735 AGCTGCGCGCGCGCGCGTCACTCCCTCGGACAGAACAGCAACAGCGCTCTTCTACGCC 1794
 QY 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAla 650
 Db 1795 GGCTCCGATCGACCTTTTACGTCAAGGACACCGGACGAGCTTC-----ACG 1845
 QY 651 AlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668
 Db 1846 CGCGGG---CCCAAGCTGGCAGCGCGGACATCCGGATATCGCTGCTCACCCGACC 1902
 QY 669 LysGluGlyAspLeuTrpLeuAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688
 Db 1903 ACCGCGGGCAGCTTGTATGTCTCGACCGAGCTCGGCATATTCGCTCCACAGACTCGGC 1962
 QY 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707
 Db 1963 ACGACCTTTGGCAGAGTCTCCACCGCCCTGACCAACACTTACAGATCGCCCTGGGTGG 2022
 QY 708 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr 727
 Db 2023 GGCTCA---GGTCTGAATCG---AACCTGTATGCTTCGGACCC-----GGCCCGTCA 2070

QY 728 GlyAla-----TyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspAsp 745
 Db 2071 GGGCTCGCCTCTAGCCAGTGGAGACAGCGCGCTCTCTGGACGACATCCAGGGCTCC 2130
 QY 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765
 Db 2131 CAGGCTTCGGCTCCATCGACAGCACCAAGGTTCGCCGCGAGCACCOCGCCGGA 2190
 QY 766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyr-----GlyAspIleGlyAla 783
 Db 2191 GTCTAGTGGGACCAACACCGCGCGCTCTTTTACGTCAGGAAACGTCGCGCGCGC 2250
 QY 784 ProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSer 803
 Db 2251 ACGGCGGGACT-----TCTCTCGACCAAGCAGAC 2283
 QY 804 ProSerProSerSerProSerProSerProSerProSerProSerProSerSerSer 823
 Db 2284 AGCAGCATGACTCTTCGCCAGCTTCGAGCACACGCTGAGGTGAGCGTTGTATCCAG 2343
 QY 824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 843
 Db 2344 ACCCGGGTTCGACGGTGACTTCGTGAGGACGACGTCGCGCGCGTCCACCGGGTCA 2403
 QY 843 rProSerPro 846
 Db 2404 GGGTTCGCG 2413
 RESULT 6
 ID ADH51580 standard; cDNA; 2710 BP.
 XX ADH51580;
 XX 25-MAR-2004 (first entry)
 XX Trichoderma reesei eg16 cDNA sequence related to ethanol production.
 DE EGV1 protein; endoglucanase; ethanol production; biomass composition;
 KW fermentation; cellulohydrolase; detergent production; softening agent;
 KW cotton fabric; wood pulp degradation; sugar; gene; ss.
 OS Hypocrea jecorina.
 FH Key Location/Qualifiers
 FT CDS 98..2557
 FT /*tag= a
 FT /product= "Mature T reesei EGV1 protein"
 FT /note= "This sequence contains numerous translation
 FT exceptions where the amino acid is given in the protein
 FT sequence as J"
 XX US2003113732-A1.
 XX 19-JUN-2003.
 XX 18-DEC-2001; 2001US-00026994.
 XX 18-DEC-2001; 2001US-00026994.
 XX (DUNN/) DUNN-COLEMAN N.
 XX (GOSD/) GOEDEGEHUR F.
 XX (WARD/) WARD M.
 XX (YAOJ/) YAO J.
 XX Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;
 XX WPI; 2004-106460/11.
 XX P-PSDB; ADH51581.
 XX New substantially purified EGV1 polypeptide with the biological activity
 XX of endoglucanase, useful for degrading biomass to ethanol.

Best Local Similarity:	40.43%	Mismatches:	293
Query Match:	30.35%	Indels:	107
DB:	10	Gaps:	26
US-09-917-376-1 (1-957) x ADD42060 (1-2849)			
QY	25	SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla	44
DB	68	TCTCAGTCTCTGCTGCTGGGGCGCTC-----ATCCCTGCC	109
QY	45	HisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGly	63
DB	110	CATGCTGCC-----TTTTCATGTGAAGACGTCAAGCTCGGGCGGGCGGCGC	157
QY	64	PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp	83
DB	158	TTTCTCCCGGCATCATCTTCATCCCAAGACAAAGGCGTAGCATATGCAACACAGAT	217
QY	84	IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrp	103
DB	218	ATTGCGGGCTGTACCGCTCAAC---GCCGACGACTCATGACCGCGTACGGATGGG	274
QY	104	Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla	119
DB	275	ATTGCTGATATGCGGGCTGGCACACTGG-----GGCATCGACGCTGTGGCTT	325
QY	120	AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp	139
DB	326	GATCCGAGGACGATCAAAAGGTGTATGCCGAGTGTATGATACGAACAGCTGGTC-	384
QY	139	-----	139
DB	385	TGTGATGTCTCTCAGATCTAGACCTATGATTGGACGCTGACATTTGGCCATATATAGGGAT	444
QY	140	ProAsnAspGlyAlaIleLeuArgSerAspGlnGlyAlaThrTrpGlnIleThrPro	159
DB	445	CCGATTAATGGAGCCATCATTCGCTCGTCAGACCGGGCGGCAACGTGGTCTTCCCAAC	504
QY	160	LeuProPheLysLeuGlyLysMetProGlyArgGlyMetGlyLysLeuAlaVal	179
DB	505	TTGCCCTTCAAAGTCCGGGGTAAATGCCAGGACGGAGCGGAGAGCGTCTGGCTGTC	564
QY	180	AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArg	199
DB	565	GATCCGGCCCACTCCCAACATCATCTTGTGTCTCGCTCAGAAACGGCTCTGGAAG	624
QY	200	SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyr	219
DB	625	TCTAGGACGGCGGTGACCTTTTCCAAAGGTCTCGTCTTCCGCAACTGGGAGCTAC	684
QY	220	IleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal	239
DB	685	ATCCAGACCCGAGTGAITCCCAACGGCTACAACAGCGACGACCAAGGACTCATGTGGTT	744
QY	240	AlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal	259
DB	745	ACGTTTCACTCAACGAGCAGCAGCCGGGGGAGCCACGTCGTATCTTTGTGGCAGC	804
QY	260	AlaAsnPro---AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla	278
DB	805	GCTGATAACATCACTGCTTCACTATGTGAGCAGCATGCGGCTCCACGTGGAGTGT	864
QY	279	ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis	298
DB	865	GTACCGGGGACCGAGGAATACTTCTCTCAAGGCGAAACTGCGAGCCAGAGAG	924
QY	299	ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerGlyAspVal	318
DB	925	GCCTTGATCTGACCTATTCGATGGCA CAGGCGCGTATGATGGCACACTTGGCTCAGTG	984
QY	319	TrpLysPheSerValThrSerGlyThrTrpArgIleSerProValProSerThrAsp	338
DB	985	TGGAGGTACGACATTGACGGGGAACTTGGAAAGACATCACCCCTCTCTCTGGATCAGAT	1044

QY	339	ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr	358
DB	1045	CTA-----TACTTTGGCTTTGGCGGCTTGGCTCGATTGTTGCAAAAGCCAGGAACC	1095
QY	359	IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp	378
DB	1096	CTTGTGTTGCTTCTTTGAACCTCTTGGTGCCAGATGCTCAGCTGTTTTCGGTCGACCGAC	1155
QY	379	GlyGlyAlaThrThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg	398
DB	1156	TCTGGACAAACATGAGCCGATCTGGCGTGGCGAGCTATCCGACTGAGACC-----	1209
QY	399	TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGln-----	414
DB	1210	TATTACTACAGATCTCAGT-GAGTCACTTTAAACGATCCGATCGCAATGGCACTGACTC	1268
QY	415	-----ProAsnProValPro-----	420
DB	1269	TGCTTCAAGACTCCCAAGACCGTGGATCAAGAACAACTTTATTCGATGTGACGAGCGAG	1328
QY	421	SerPro-----LysLeuGlyTyrMetAspGluAlaMetAlaIleAsp	434
DB	1329	TCACCGTCCGATGGTCTCATCAAGCGCTCGGCTGGATGATTGAGTCTCTCGAGATTGAC	1388
QY	435	ProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAsp	454
DB	1389	CAACCGCAGCAGCAACCATCTGGCTCTACGGCACCGGATGACATCTTTGGCGGCCACGAT	1448
QY	455	LeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu	474
DB	1449	CTCACCACTGGGACACGCGCCACAAATGTCAATCAATCACTGTGGCAGACGGCATCGAG	1508
QY	475	GluThrAlaValAsnAspLeulleSerProProSerGlyAlaProLeulleSerAlaLeu	494
DB	1509	GAATTCCTCGCTCAGGACCTCTGACCCCGCGGGAAGCAGGCTATTGGCGCGATC	1568
QY	495	GlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSerThr	511
DB	1569	GGAGACGACAAACGGCTTTCACCTTTCAGCAGAAACGACCTCGGGACATCGCCGACAGC	1628
QY	512	IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnPro	531
DB	1629	GTCCTGGCAACGCCCATGCGGCACCTCGACGAGCTCGACTACGCGCGGAACCTCGTC	1688
QY	532	SerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisVal	551
DB	1689	AAGAGCGTCTGCTCGCGTCCGCAACACCGCGGCAACGCAACAG-----GTG	1733
QY	552	AlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThr	571
DB	1734	GCAATCTCGTCCGACCGCGCGCGAGCTGAGCATCGACTACGCGGCGGACAGCTCCATG	1793
QY	572	ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAsp	591
DB	1794	AACGGCGGACCGTGGCTATTTCGGCGGCGGCGACGATCTCTCTGTCGACCGCGCTCG	1853
QY	592	ProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGly	611
DB	1854	TCCGGC-----GTGACGCGCTCGCAGTTCAGGGCAGCTTTCGCTCGTCCGAGC	1904
QY	612	ValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeu	631
DB	1905	CTGCCCGGGCGCGCTCATCTCGCTCGGACAAAGAACCAACAGCGCTCTCTACGCGCGC	1964
QY	632	SerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAla	651
DB	1965	TCCGGATCGACCTTTTACGTACCAAGGACACCGGACGAGCTTC-----ACGCGC	2015
QY	652	GlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGlyLys	669
DB	2016	GGG---CCCAAGCTGGGCGGCGGACGATCGGGGATATCGTGTCTCACCCGACCAACC	2072


```
QY 670 GluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySer 689
Db 2073 GCGGCGAGTGTATGTCGACCGACGTCGTCATATTCGGCTCCACAGATCGGCGCAG 2132
QY 690 SerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 708
Db 2133 ACCTTTGCCCAAGTCTCCACCGCCCTGACCAACACCTACCATCGCTCGGTGTGGGC 2192
QY 709 AlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyValThrGly 728
Db 2193 TCA---GGCTCGAATGG---AACCTGTATGCTCTTCGGCACC-----GGCCCGTCAGG 2240
QY 729 Ala-----TyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGln 746
Db 2241 GCTCGCTCTACCGCAGTGGAGACAGCGCGCTCTCTGGACGGACATCCAGGGCTCCAG 2300
QY 747 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 766
Db 2301 GGCTTCGGCTCCATCGACAGCACCAAGTTCGCCGCGCAGCGGCGACCGCGCGCAAGTC 2360
QY 767 TyrIleGlyThrAsnGlyArgGlyIleValTyr-----GlyAspIleGlyAlaPro 784
Db 2361 TACGTGGGCACCAACAGCGCGGCGCTCTTTTACGCTCAGGGAAACGCTCGCGCGGCACG 2420
QY 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804
Db 2421 GCGCGGACT-----TCTCGTCCACCAAGCAGACGACG 2453
QY 805 SerProSerSerSerProSerProSerProSerProSerProSerProSerProSerPro 824
Db 2454 AGCAGTACTCTTCGCCAGCTCGAGACACACGCTGAGTGGAGCTGTATCCACGACG 2513
QY 825 SerProSerProSerProSerPro---SerProSerArgSerProSerProSerProSerPr 844
Db 2514 CGGGCTTCGACGGTACTTCGTTCGAGGACACGCTCGCGCGCGCTCCACCGGGGTTCAGGG 2573
QY 844 cSerPro 846
Db 2574 GTCGCG 2580

RESULT 8
ID ADD24893 standard; DNA; 3668 BP.
XX AC ADD24893;
XX 15-JAN-2004 (first entry)
XX DE DNA encoding Xanthomonas campestris cellulase #1.
XX KW Directed genetic engineering; galactomannanase; reduced activity;
XX KW enhanced activity; xanthan gum production; suspension stability;
XX KW emulsion stability; temperature resistance; pseudoplasticity; amylase;
XX KW cellulase; extracellular protease; intracellular protease;
XX KW glucose dehydrogenase; gene; ds.
XX OS Xanthomonas campestris.
XX PN US2003036176-A1.
XX XX 20-FEB-2003.
XX PF 10-AUG-2001; 2001US-00927827.
XX XX 28-MAR-2001; 2001US-0279493P.
XX XX (BOWE/) BOWER S G.
XX PA (RAMS/) RAMSEIER T M.
XX XX Bower SG, Ramseier TM;
XX DR WPI; 2003-625389/59.
XX DR F-PSDB; ADD24919.
```

```
XX New transformed cell or organism having reduced or enhanced activity of
PT at least one protein, useful for producing xanthan gum, which are useful
PT for providing formulations and properties.
XX Claim 15; SEQ ID NO 21; 135pp; English.
XX The present invention relates to polypeptide and polynucleotide sequences
CC from Xanthomonas campestris which may be used for activity reduction or
CC enhancement using directed genetic engineering. A transformed cell or
CC organism having reduced or enhanced activity of at least one such protein
CC e.g. galactomannanase can be generated by disrupting the gene encoding
CC the protein. The activity of the protein is reduced by the presence of an
CC antisense nucleic acid sequence. The nucleic acid sequence of the gene
CC encoding the protein is a recombinant sequence having at least one
CC mutation as compared to the wild-type gene encoding the protein. The
CC transgenic cell or microorganism are useful for producing xanthan gum,
CC which are useful for providing formulations and properties, such as long-
CC term suspension and emulsion stability in alkaline, acid, and salt
CC solutions, temperature resistance, and pseudoplasticity. The present
CC sequence encodes an enzyme relating to the present invention.
XX SQ Sequence 3668 BP; 570 A; 1181 C; 1292 G; 625 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9, 11e-33 Length: 3668
Score: 1167.50 Matches: 299
Percent Similarity: 48.53% Conservative: 137
Best Local Similarity: 33.30% Mismatches: 351
Query Match: 22.74% Indels: 112
DB: 10 Gaps: 28

US-09-917-376-1 (1-957) x ADD24893 (1-3668)
QY 23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42
Db 1136 ACCACACAGGAGCATGGCTGCTGTCTTGTAGCCTGTG---CTGTGTCTCTTTTCGACG 1192
QY 43 Pro-----AlaHisAlaIleThrGlnProTyrThrTrpSerAsnValAla 58
Db 1193 CCCAGCTGCGTGGCGGCGAGCCGCCAGCTCCGCGCCCTACGAGTGGCGCATGTGCGCC 1252
QY 59 IleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78
Db 1253 AITGGCGTGGCGGCTTTGTACCGGTGTCTTTTCATCCCGCGCAAGCTGTGTGGCC 1312
QY 79 TyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle 98
Db 1313 TATGCGCGCACCGATGTGGTGGCGGTACCGTGGGATGCGCAGCGCAGCATGGACC 1372
QY 99 ProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118
Db 1373 GCGCTGACCGACTGCTGGCGCTGACGACTGGAACCTGATGGCATCGACGCAATTTCGCC 1432
QY 119 AlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 138
Db 1433 GTCGACCCCGCGCATGCCGATGCTATCTGGCGCGCGGCACCTATATGCAAGAA--- 1489
QY 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaIleThrTrpGlnIleThr 158
Db 1490 CGCGCGGCACTGCGCAGTGTGGCTGCTTCACCGCGCGCGCATCGTTCGAGCGTGC 1549
QY 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla 178
Db 1550 GACCTGCGCTTTAAGCTGGTGGTAACACAGCTGGCGCGCCCAATGCGCAGCGCTGGCG 1609
QY 179 ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198
Db 1610 GTGGACCCCGCACGATGGCGCGCTGTCTGTCTGGGTCTCGCGCATGCC---GGCCTGTGG 1666
QY 199 ArgSerThrAspSerGlyAlaIleThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218
Db 1667 CGTAGCAGGATCGCGCGCGCACCTGGGCGAAGGTGGCGTCTGTTTCCGACCGCGCGCTG 1726
```


Qy 219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyr 238
Db 1727 GCCGGTCCACCGCGCGCAATCATGTTGGCGGCGAGCAGCGGGT---GGATCGCCTTT 1783
Qy 239 ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly 258
Db 1784 GTCTGTTTCAGCGAGCAGCGGTGGCAACACGGCTCGCCACACCGCGCATCTAGTGGG 1843
Qy 259 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 278
Db 1844 GTGTCCACCGAACAGACAGCAGCTGTATGTGTCCGAAGATGCGCGCGCAGTTGGCCACCG 1903
Qy 279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
Db 1904 GTGCCGGGCAACACCGCGCTCGCGCCGAGCCACATGCGCGCGCGAGCGATGGGCAC 1963
Qy 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318
Db 1964 ---TGGTATCTGAGCTATGGCGACACAGCCCGGCGCGACCTGATGCGGGGAGCCTTG 2020
Qy 319 TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338
Db 2021 TGGAAATTCAGCGCGCACAGGGCGCTGGCGTGAGATCAGCCCGATTCCGCAG---CCA 2077
Qy 339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
Db 2078 GCCAGTGGCATGATTCGGCTGGGTGCGGTGGGTGGATCGGCAACATCCGACGGTG 2137
Qy 359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIlePheArgSerThrAsp 378
Db 2138 CTGCTGGCGACACCTTCGCGCGTGGCGCGCGCGAGAGCTGTATCGCAGCGTGGAT 2197
Qy 379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398
Db 2198 GTGGCAAGCACTGGACCGCTG-----TTGGCCGAT 2230
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProPro 418
Db 2231 GCGGTGTTTCGATCACAGCGCGCGCGCTG---ACCGCACATGCCACCGCGCAC----- 2281
Qy 419 ValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSer 438
Db 2282 -----TGGATGGG---GCGCTGGCGATCGATCCGTTTCAGCGC 2317
Qy 439 AspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrp 458
Db 2318 AACCATGCGCTGTTCTGACCGGCTACCGCATCTGGGCTCGCGCAATCTG----- 2368
Qy 459 AspSerGlyGlyGlnIleAlaPro-----MetVal 470
Db 2369 -----CAGATTTCGCGCACCGCAGCGCGCTGCGAGTGGTGTTCAGGAC 2416
Qy 471 LysGlyLeuGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu 490
Db 2417 CGTGGCTTGGAAAGAACCGTGGCTGGACCTGTCTCAGCCCGATGGTGGCGCGCATCTG 2476
Qy 491 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 510
Db 2477 CTCAGCGCTGCGGATACCGCTTCGGGCATGACGACCTGGACCGGCTG---CAG 2533
Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
Db 2534 TTGCAGTACCGCGGCCACCGCTGACCAATGGCGAAAGATCATGATGCGCGCGCAGCGG 2593
Qy 531 ProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550
Db 2594 CCGCAGTGGGTGGCGAGCGGTACCGTGGCGCGCGCGCGCAACAGAAATCCGCGCG 2653
Qy 550 sValAlaPheSerThrAspGlyGlyAsnTrpPheGln---GlySerGluProGlyG 569
Db 2654 C---TGTATTACGGCTGATGGCGGCAAGCAATGGAGCTCGGTTTGGAGTGGAGCCCGCAGC 2710

Qy 569 yValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPr 589
Db 2711 CGGCGAGCGCGCGCGAGCATTCATGGTGGCGGATCGCGAGGTGGTGTGGCGACC 2770
Qy 589 oGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSe 609
Db 2771 GGCGCGCGCGGCAAT-----TGGCGCACGTCGACTTCGGCGCGCATGTGCAGCGTGT 2824
Qy 609 xGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTy 629
Db 2825 GAACGCCCTTCGCGAACCCCGGTGGTGGTGGATCGGTGGGAGCAGCGCGCTGGTA 2884
Qy 629 rAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGl 647
Db 2885 TGGCGTGGATGTCGCCAGCGGCGAGCTGTACGAGAGCACCGATGCGCGCGCAGTTCCG 2944
Qy 647 nProValAlaAlaGlyLeu----- 653
Db 2945 T-----GCGACCGGTGTACAGGTGGCGAGCCCGCGCGCATGAGCGCACCGCGCGCA 2998
Qy 654 -----ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGl 671
Db 2999 GCTCGTTCGAGCCCGCGCGCGCGGTGTGTATCTGGCCAGCCCGCGCAAG----- 3053
Qy 671 yAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTr 691
Db 3054 -----GGTGTGTGTGGTGGCAGCGGT-----CGGCT 3082
Qy 691 pSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGl 711
Db 3083 CGAGTGTCTCTCACAGCCGCGAGACAGCTCGTGGGCATCGCGAGCGGTGGCGG 3142
Qy 711 ySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrAr 731
Db 3143 CGGTGGCGCGCGCGCTGTATCTGGCGCGCGGTGGAGCGGTGGTGTTCG 3202
Qy 731 sSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAs 751
Db 3203 CTCGACCATGGCGCGCGCGAGTGGCGAGCATCAAGATGACGCGCACCGTTCGGCGG 3262
Qy 751 nTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAs 771
Db 3263 C---CCGTACAGCGTGCAGCGGTGATCCGCGCATTCGCGCGCTGTGTACTTCGCCACCGG 3319
Qy 771 nGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSe 791
Db 3320 CGCGCGCGCATTTTCTACGCGCATCCGAGATGAGCAGTGCATGCCCTGTCCGCGA----- 3374
Qy 791 rValSerProSerAlaSerProSerLeuSerProSerProSerProSerProSerPro-- 810
Db 3375 ---CGTCTGTGTCTTACCGTGTACGCGCGCGCGCGCTGGCCCTGCGGGCGCGGTG 3430
Qy 811 -----SerProSerProSerProSerProSerProSerProSerProSerPro-- 827
Db 3431 ACACGCGCGACCTCGAGATCACCGCATCCCGCGT---CCACATTTCCGCGCGCTGTGTTG 3489
Qy 828 -----ProSerProSerProSerProSerArg----- 836
Db 3490 CTGGGTGTGTGCTGGCGAGCGCGCGCGCGCGCAACCGAGCGCTGCGGTGTGCTGTTT 3549
Qy 837 -----SerProSerProSerAlaSerProSerProSerProSerProSerPr 850
Db 3550 GCCGATGGCGGTGTGTCAGCGCATGAGGCAATGCCGGTATGGGGCTGGGCGCACGCCC 3609
Qy 850 oSerProSerSer-----SerProSerSerProSerProThrPro 864
Db 3610 GCGCGCGCAATCACGGTGTGTGATGGCAAGCAGCGCCACCGCGCAAGACCG 3661

RESULT 9
ADH19117
ID ADH19117 standard; cDNA; 2646 BP.
XX ADH19117;
AC

11-MAR-2004 (first entry)
 Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme precursor cDNA.
 xyloglucan oligosaccharide degradation; genetic engineering; ss; gene;
 precursor.
 Geotrichum sp. M128.
 Key Location/Qualifiers
 CDS 120..2558
 /*tag= a
 /product= "Geotrichum sp M128 xyloglucan oligosaccharide-
 degrading enzyme precursor protein"
 EP1350844-A2.
 08-OCT-2003.
 25-MAR-2003; 2003EP-00251866.
 25-MAR-2002; 2002JP-00083433.
 (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 Yaoli K, Mitsuishi Y;
 WPI; 2004-100948/11.
 P-PSDB; ADH19118.
 Novel xyloglucan oligosaccharide-degrading enzyme with a different
 degradation mechanism from known enzymes.
 Claim 11; SEQ ID NO 11; 39pp; English.
 The invention relates to a novel xyloglucan oligosaccharide-degrading
 enzyme with a different degradation mechanism from known enzymes. The
 primary structure and polynucleotide structure of the xyloglucan
 oligosaccharide-degrading enzyme provided by the present invention allows
 high-purity polypeptides having a xyloglucan oligosaccharide-degrading
 activity to be prepared at a low cost through a genetic engineering
 process. The current sequence is that of the Geotrichum sp. M128
 xyloglucan oligosaccharide-degrading enzyme precursor cDNA of the
 invention.
 SQ Sequence 2646 BP; 550 A; 866 C; 787 G; 443 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2.09e-31 Length: 2646
 Score: 1125.50 Matches: 287
 Percent Similarity: 49.82% Conservative: 131
 Best Local Similarity: 34.21% Mismatches: 294
 Query Match: 21.92% Indels: 127
 DB: 12 Gaps: 32
 US-09-917-376-1 (1-957) x ADH19117 (1-2646)
 QY 17 ValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuPro 36
 DB 114 TTAGTATGGTGGCAGTCACTCCCTCGGGAAGGGCTTACTGCTTTCGATTCG--- 170
 QY 37 IleAlaIleThrAlaSerProAlaHisAlaAlaThrGlnProTyrThrTrpSerAsn 56
 DB 171 -----GCGTCGTCGCGGTGCGCAAGGAGCACTACGAGTTCGAAGAT 212
 QY 57 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76
 DB 213 GTCGCGATCGCGCGGGGTACATTACCGGGGATTTGTCGGCGACCCCAAGACCAAGGAC 272
 QY 77 IleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 96
 DB 273 CTGCTGTACGCGCGCACGACATTGGCGGCGCGTACCGCTGGGAGCGGACGTCCTCAAG 332

QY 97 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSer 116
 DB 333 TGGATCCCGCTCAACGACTTTATCGAGCGCGAGACATGAACATTATGGCACCCGAGTCG 392
 QY 117 IleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThr--- 135
 DB 393 ATCGCGCTGGACCCCAACACCCGACAGGCTGTACTCTCGGCGACGGGCGGTATGTTCGCGC 452
 QY 136 AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrp 155
 DB 453 GACGAGTGG-----GCGCGCTTCATGTGTCGGAAGACCGCGGCGCAGTCGTTT 500
 QY 156 GlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlu 175
 DB 501 ACAATCTACGAGTCCCGGTTCCGATGGGCGCCACGACATGGGACGCAACAATGGCGAG 560
 QY 176 ArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys 195
 DB 561 CGCCTCGCTGTCACCCGTTCAACTCGAACGAGGTCGTGATGGGTACGCGTACA---GAG 617
 QY 196 GlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAsp 215
 DB 618 GGTATCTGGAAGAGTTCGACCGCGCCCAAGACCTGGACAAACGTCACGTCCATCCCGGAC 677
 QY 216 ValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGly 235
 DB 678 GCGTTC-----ACCAACGGTATCGGATACACGTCG----- 707
 QY 236 ValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIle 255
 DB 708 -----GTCATTTTCGACCCC-----GAACGTAATGGCACCATC 740
 QY 256 PheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyAlaThr 275
 DB 741 TACGCGAGCGGCGACTGCGCCCGCGAGGCG---ATGTACGTCACGACGACGCGGTCTCG 797
 QY 276 TrpGlnAlaValProGlyAlaProThrGlyPheIle----- 287
 DB 798 TGGGAGCGAGTGGCGGCGGCGAGCCGTCAGCTGGCTCAACAGGACACACGGGCGGTTCGCG 857
 QY 288 -----ProHisLysGlyValPheAspProValAsn 297
 DB 858 GACAAGAGCGCGCTCGATCGCGCGGAGCCCATGAAAGTCGCTCTCACCCCC----- 911
 QY 298 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAsp 317
 DB 912 AACTTCCTCTACTGACTTACGCCGACTACCTCGGTCTCCATGGGCGGTCTACGTTTCGCGAA 971
 QY 318 ValTrpLysPheSerValThrSerGlyThrTrpThrArgIle----- 331
 DB 972 GTCTGGCGCCAGAACCGACCTCGGGCGCTGGGACGACATTATCTCCCGCGTCGCGCAAC 1031
 QY 332 --SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 350
 DB 1032 TCGTCGCTCTCGCCCGTACAAACACGAGTTCCTCGGGCGGATTTTCGGTCTCAGC 1091
 QY 351 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 370
 DB 1092 GTCGAGCGGACCAACCCCAACCGTCTCGTCTCATCACCC---CTCGACCGCGACCCCGGA 1148
 QY 371 ThrIle-----IlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387
 DB 1149 CCCGCCCTCGACGATCTACTCTCAACCCGATGCGGCGCGACCTGGGAAGGACGTCACC 1208
 QY 388 AspTrpThrSer-----TyrProAsnArgSerLeuArg 398
 DB 1209 CAGCTCTCTCCCTCCCTCCAACTCGAAGTAACTGGGGCCACCCTGACTAACCGGCGCGG 1268
 QY 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPro----- 415
 DB 1269 TAC---AAGGACGGCACGCTGTTCCGTGGCTCGACTTCAACACGCGTCCCCCAGTGGGG 1325

QY 416 -----AsnProValProSerPro-----LysLeuGlyTyrMetAspGluAla 430
 Db 1326 GGATACGGTGGCGCGACGAGTACCGCGCGCTCACAAAGTTGGCTGGTGGATGAGCGCT 1385
 QY 431 MetAlaIleAspProPheAsnSerArgMetLeuTyrGlyThrGlyValAlaThrLeuTyr 450
 Db 1386 GTGCTTATCGATCGTTCAACCCGACGACCTGATGATCGGACGCGGGCGGACCATCTGG 1445
 QY 451 AlaThrAsnAspLeuThrLys-----TrpAspSerGlyGlyGlnIleHisIle 466
 Db 1446 GCGACCGACACGCTCTCCCGTGTGAGAAGGACTGG----- 1481
 QY 467 AlaPro-----MetValLysGlyLeuGluThrAlaValAsnAspLeuIle 482
 Db 1482 GCGCGAGCTGGTACCTCCAGATCGACGATTCGAGGAGATCGCATCTCTGCTCGTCCGC 1541
 QY 483 SerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 502
 Db 1542 TCGCCCAAGAGCGCGCGCGCTCTGTCGGCATCGGTGATAGCGGCATGAGACAC 1601
 QY 503 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr 522
 Db 1602 GAGGACCTCACCAAG-----CCGACAGAGATGTTGTGGCGCCCGCTCTCCAACTCGAC 1658
 QY 523 SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 542
 Db 1659 AGCATCGACGCTGCGGCAACTTCCCAACGTTGTGTCGCGCGCGGATCTCTCGGACAC 1718
 QY 543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp--- 561
 Db 1719 GAGTACGACAGCGCGTGGCGCGCGTGGTACGCGCATGACGCGGAGACCGCTGGACC 1778
 QY 562 ---PheGlnGlySerGluProGlyGlyValThrThr-----GlyGlyThrValAla 577
 Db 1779 ATCTTCCCTACTGCTCTGCGCATGAACGCGGACCTACCGGCGGACGACGATGCA 1838
 QY 578 AlaSerAlaAspGlySerArgPheValTrpAlaPro-----GlyAspProGlyGlnPro 595
 Db 1839 GTCGCGGCGGCGGCGGACGATCGTGTGTCGACCAAGCTTGACGACGAGCGCTCGGA 1898
 QY 596 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla--- 614
 Db 1899 CCGTGTACTCGACGACTATGCAAGACGTGCT-----GTTCCCGCTGGC 1946
 QY 615 -----AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629
 Db 1947 GACCTGAGCGCCACACTGCCAATGTGCTCTCGACAGGTCCAGATGCGACGTTCTAC 2006
 QY 630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649
 Db 2007 GCTACCGATGGCGCAAGTTCTCGTCTCGACCGCGGGAAGTCGTATGCGCGCAAG 2066
 QY 650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 669
 Db 2067 GCGCGCGGACTTGTCACT-----GECACATCGCTCATCGCTCGCGTGAACCCCTGGTG 2120
 QY 670 GluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688
 Db 2121 GCGCGGAGCTGTGGTGCTGTTCGAGGGCGGTCTCTTCCACTCGACCGACTTGGC 2180
 QY 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707
 Db 2181 GCCTGTTCACGAGGTAGGTACCGCAACCGCACCTCGTGTGAGCGTGGCGGCCCAAG 2240
 QY 708 Ser-----AlaProGlySerSerTyrProAlaValPheValGlyThr--- 722
 Db 2241 TCCAGTCCGACGCGCAAGAGCTAGCGCGCTTCCGCGGTCTTCATCTGGGACCGAC 2300
 QY 723 ---IleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeu 741
 Db 2301 AAGCTTGAAGGACATCGGCTGTACCGCTTCGACGACACGCGGACCTGGAGCGCGC 2360
 QY 742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761

Db 2361 GTCAATGACGAGGACACAACTACTCTCGGCG---CCACCATGATCGAGCGGACCCCAAG 2417
 QY 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
 Db 2418 GTCTACGGCGCGTGTATCTAGGACGAACGCGCGGATCTGTGTACGCCGACCTT 2474
 RESULT 10
 ADH19123
 ID ADH19123 standard; DNA; 2481 BP.
 XX
 AC ADH19123;
 XX
 DT 11-MAR-2004 (first entry)
 XX
 DE Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme-derived DNA.
 XX
 KW xyloglucan oligosaccharide degradation; genetic engineering; ds; gene;
 XX His tag.
 OS Synthetic.
 OS Geotrichum sp. M128.
 FH Key Location/Qualifiers
 CDS 1..2481
 FT /*tag= a
 FT /product= "Geotrichum sp M128 xyloglucan oligosaccharide-
 FT degrading enzyme-derived protein"
 XX
 PN EP1350844-A2.
 XX
 PD 08-OCT-2003.
 XX
 PF 25-MAR-2003; 2003EP-00251866.
 XX
 PR 25-MAR-2002; 2002JP-00083433.
 XX
 PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX
 PI Yaoi K, Mitsuishi Y;
 XX
 DR WPI: 2004-100948/11.
 DR P-PSDB; ADH19124.
 XX
 PT Novel xyloglucan oligosaccharide-degrading enzyme with a different
 PT degradation mechanism from known enzymes.
 PS Example 7; SEQ ID NO 17; 39pp; English.
 XX
 CC The invention relates to a novel xyloglucan oligosaccharide-degrading
 CC enzyme with a different degradation mechanism from known enzymes. The
 CC primary structure and polynucleotide structure of the xyloglucan
 CC oligosaccharide-degrading enzyme provided by the present invention allows
 CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation
 CC activity to be prepared at a low cost through a genetic engineering
 CC process. The current sequence is that of the Geotrichum sp. M128
 CC xyloglucan oligosaccharide-degrading enzyme-derived DNA of the invention
 CC which comprises the wild-type enzyme sequence plus a C-terminal region
 CC including a Histidine tag originating from vector pET29a(+).
 XX
 SQ Sequence 2481 BP; 496 A; 836 C; 748 G; 401 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,24e-31 Length: 2481
 Score: 1124.00 Matches: 298
 Percent Similarity: 48.91% Conservative: 127
 Best Local Similarity: 34.29% Mismatches: 306
 Query Match: 21.89% Indels: 138
 DB: 12 Gaps: 33

US-09-917-376-1 (1-957) x ADH19123 (1-2481)

QY 52 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71

1009	TCGAGGACGCTCACCACGCTCTCGTCCCGTCCAACCTCGAAGGTAACTGGGGCCACCG	1061	TCGAGGACGCTCACCACGCTCTCGTCCCGTCCAACCTCGAAGGTAACTGGGGCCACCG
394	AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpIleThrPheGlyVal	413	AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpIleThrPheGlyVal
1069	ACTAAGCGCGCGGTAC---AAGGACGGCAGCGCTGTTCCGTGGCTCGACTTCAACAAC	1125	ACTAAGCGCGCGGTAC---AAGGACGGCAGCGCTGTTCCGTGGCTCGACTTCAACAAC
414	GlnPro-----AsnProProValProSerPro-----LysLeuGly	425	GlnPro-----AsnProProValProSerPro-----LysLeuGly
1126	GGTCCCGCATGGGGGGGATACGGTGGCGCGCAGCGGTACGCCCGGCTCACCAAGTTTGGC	1185	GGTCCCGCATGGGGGGGATACGGTGGCGCGCAGCGGTACGCCCGGCTCACCAAGTTTGGC
426	TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr	445	TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr
1186	TGGTGATGACGCGCTGTGCTTATCGATCGTTCATCCCGGACGACCTGATGATCGGCACG	1245	TGGTGATGACGCGCTGTGCTTATCGATCGTTCATCCCGGACGACCTGATGATCGGCACG
446	GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly	461	GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly
1246	GGGGCGACCATCTGGGCGACCGACGACGCTCTCCGTGTCGAGAAGGACTGG-----	1296	GGGGCGACCATCTGGGCGACCGACGACGCTCTCCGTGTCGAGAAGGACTGG-----
462	GlyGlnIleHisIleAlaPro-----MetValIysGlyLeuGluThrAla	477	GlyGlnIleHisIleAlaPro-----MetValIysGlyLeuGluThrAla
1297	-----GCGCGAGCTGGTACCTCCAGATCGACGGTATCGAGGAGAAATGCG	1341	-----GCGCGAGCTGGTACCTCCAGATCGACGGTATCGAGGAGAAATGCG
478	ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu	497	ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu
1342	ATCTGTCGCTCGCTCGCCCAAGAGCGCGCGGCGCTCTGTCTGGGCATCGGTGACATT	1401	ATCTGTCGCTCGCTCGCCCAAGAGCGCGCGGCGCTCTGTCTGGGCATCGGTGACATT
498	GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal	517	GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal
1402	ACGGCATGAAGCAGCAGCAGCTCACCAAG---CCCCAGAAGATGTTTGGTGGCCCCAG	1458	ACGGCATGAAGCAGCAGCAGCTCACCAAG---CCCCAGAAGATGTTTGGTGGCCCCAG
518	PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla	537	PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla
1459	TTTCTCAACTCGACAGCATCGCTCGGGCAACTTCTCCCAACGTTGTCTGCTCCGCC	1518	TTTCTCAACTCGACAGCATCGCTCGGGCAACTTCTCCCAACGTTGTCTGCTCCGCC
538	GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly	557	GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly
1519	GGATCTCGGACACAGTACACAGCGCGTGGCGCGGTGCTGTCGCGACTGACGCGC	1578	GGATCTCGGACACAGTACACAGCGCGTGGCGCGGTGCTGTCGCGACTGACGCGC
558	GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThr	572	GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThr
1579	GGAGACGCTGGACCATCTTCCCTACCTCCCTCTCGCATGAACGCGACCATACCAG	1638	GGAGACGCTGGACCATCTTCCCTACCTCCCTCTCGCATGAACGCGACCATACCAG
573	GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly	590	GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly
1639	GGCAGCACCATTTGCGATCGACGCGCGCAGCAGATCGTGTGTCGACCAAGCTTGAC	1698	GGCAGCACCATTTGCGATCGACGCGCGCAGCAGATCGTGTGTCGACCAAGCTTGAC
591	AspProGlyGlnProValValThrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln	610	AspProGlyGlnProValValThrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln
1699	GAGCAGCCTCGGGACCGTGTGCTCGACGACTATGGCAAGCTGGTCT-----	1749	GAGCAGCCTCGGGACCGTGTGCTCGACGACTATGGCAAGCTGGTCT-----
611	GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn	624	GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn
1750	---GTTCCCGCTGGCGACCTGAAGGCCAGACTGCGCAATGTCTCTCGACAAAGTCCAG	1806	---GTTCCCGCTGGCGACCTGAAGGCCAGACTGCGCAATGTCTCTCGACAAAGTCCAG
625	ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyVal	644	ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyVal
1807	GATGGCACCTTCTACGCTACCGATGGCGGCAAGTTCTTCTCTCGACCGACGCGGGAAG	1866	GATGGCACCTTCTACGCTACCGATGGCGGCAAGTTCTTCTCTCGACCGACGCGGGAAG
645	ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis	664	ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis
1867	TGATATGCGCCCAAGGGCGCGGACTTGTACAT-----GGCACATCGCTCATGCCTGCC	1920	TGATATGCGCCCAAGGGCGCGGACTTGTACAT-----GGCACATCGCTCATGCCTGCC
665	AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis	683	AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis
1921	GTGAACCCCTGGGTGGCGGCGAGCTGTGGTGCCTTGTTCGAGGGCGGTCTCTCCAC	1980	GTGAACCCCTGGGTGGCGGCGAGCTGTGGTGCCTTGTTCGAGGGCGGTCTCTCCAC
684	SerThrAsnGlyLysSerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn	702	SerThrAsnGlyLysSerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn
1981	TCGACCGACTTTGGCGCTCTTACGAGGGTAGGTACCGCAACGCGACCTCTGTCGAGC	2040	TCGACCGACTTTGGCGCTCTTACGAGGGTAGGTACCGCAACGCGACCTCTGTCGAGC
703	ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe	718	ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe
2041	GTCCGGCCCCCAAGTCCAAAGTCGGACGCAAGAGGCTAGCGCCCTCCCGGGTCTTC	2100	GTCCGGCCCCCAAGTCCAAAGTCGGACGCAAGAGGCTAGCGCCCTCCCGGGTCTTC


```
QY 719 ValValGlyThr-----IleGlyValThrGlyAlaTyrArgSerAspCysGly 736
DB 2101 ATCTGGGGACCGCAACAGCTGGAAAGCGACATCGGCCTGTACCGCTCCGACGACACGCG 2160
QY 737 ThrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756
DB 2161 AGCACTCGACGCGGTCAATGACGAGGACCACTACTCGGC---CCACCATGATC 2217
QY 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776
DB 2218 GAGGCGGACCCCAAGGTCTACGGCGCGTGTATCTAGGCACCAACGCGCGGTATCGTG 2277
QY 777 TyrGlyAspIleGlyAlaProSerGlySerProSerProSerProSerValSerProSerAla 796
DB 2278 TACGCGGACCTTACCACCAAGAAGAGC---AACGAGGAGAAAGTCGACCCCAAGGTGCGGC 2334
QY 797 Ser-----ProSerLeuSer 801
DB 2335 AACGGCCGAGGCGACGACTGTATGTGMAAAGGAGATCTGGGTACCTCGTGCGCAC 2394
QY 802 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSe 821
DB 2395 GCGGTTCAATGGCTGATATCGGATCGAATTCGAGCTCGTCGACACAGCTTCGCGCGCA 2454
QY 821 rSerSer-ProSerProSerPro 828
DB 2455 CTCGAGCACCAACCAACCA 2477

RESULT 11
ADH19119
ID ADH19119 standard; DNA; 2367 BP.
AC ADH19119;
XX 11-MAR-2004 (first entry)
DE Geotrichum sp M128 xyloglucan oligosaccharide-degrading enzyme mat DNA.
KW xyloglucan oligosaccharide degradation; genetic engineering; ds; gene;
KW mature.
XX Geotrichum sp. M128.
XX Key Location/Qualifiers
FH 1. 2367
FT /partial
FT /tag= a
FT /product= "Geotrichum sp M128 xyloglucan oligosaccharide-
FT degrading enzyme mature protein"
FT /notes= "No start or stop codon"
XX
XX EP1350844-A2.
XX
XX 08-OCT-2003.
XX
XX 25-MAR-2003; 2003EP-00251866.
XX
XX 25-MAR-2002; 2002JP-00083433.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX
XX Yaei K, Mitsuishi Y;
XX
XX WPI; 2004-100948/11.
XX P-PSDB; ADH19120.
XX
XX Novel xyloglucan oligosaccharide-degrading enzyme with a different
XX degradation mechanism from known enzymes.
XX
XX Claim 8; SEQ ID NO 13; 39pp; English.
XX
XX The invention relates to a novel xyloglucan oligosaccharide-degrading
```

```
CC enzyme with a different degradation mechanism from known enzymes. The
CC primary structure and polynucleotide structure of the xyloglucan
CC oligosaccharide-degrading enzyme provided by the present invention allows
CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation
CC activity to be prepared at a low cost through a genetic engineering
CC process. The current sequence is that of the Geotrichum sp. M128
XX xyloglucan oligosaccharide-degrading enzyme mature DNA of the invention.
SQ Sequence 2367 BP; 471 A; 797 C; 719 G; 380 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,79e-31 Length: 2367
Score: 1117.00 Matches: 281
Percent Similarity: 50.37% Conservative: 124
Best Local Similarity: 34.95% Mismatches: 279
Query Match: 21.75% Indels: 120
DB: 12 Gaps: 31
```

US-09-917-376-1 (1-957) x ADH19119 (1-2367)

```
QY 52 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71
DB 10 TACGAGTTCAAGATTCGCGATCGCGCGGGGTACATTACCGGGATGTGCGGCAC 69
QY 72 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 91
DB 70 CCAAGACCAAGACCTGTGTACGCGCGACGACATTTGGCGCGGTACCGTGGAC 129
QY 92 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 111
DB 130 GCAGGCACGCTCCAAATGGATCCGCTCAACGACTTTATCGAGCGCGACATCAACATT 189
QY 112 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131
DB 190 ATGGGCACCGAGTCGATCGCTGGACCCCAACACCCGACAGGTGTCTCTCGCGAG 249
QY 132 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 150
DB 250 GGGCGCTATGTGCGCGACAGTGG-----GCGCGTCTCTATGTCTCGAGAC 297
QY 151 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 170
DB 298 CGCGGCGAGTCGTTTACATCTACGAGTCGCGCTTCCCGATGGCGGCAACGACATGGGA 357
QY 171 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 190
DB 358 CGCAACAATGGCGAGCGCTCGCTGTCAACCCGTTCACTCGAACGAGGTCTGGATGGGT 417
QY 191 AlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMet 210
DB 418 ACGCGTACA---GAGGGTATCTGGAAGAGTTCGAGCCGCCCAAGACTTGGACAAACGTC 474
QY 211 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 230
DB 475 ACGTCCATCCGACCGGTTTC-----ACCAACGGTATCGGATACACG 516
QY 231 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGln 250
DB 517 TCG-----GTCAATTTGACCC-----GAA 537
QY 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 270
DB 538 CGTAATGGCACCATCTACGCGAGCGGACTGCCCCCGCAGGGGC---ATGTCGTCCGAC 594
QY 271 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProGlyAlaPheIle----- 287
DB 595 GACGCGGTGTCTCGTGGAGCCAGTGGCGGCGCAGCGTCCAGCTGGCTCAACAGGACC 654
QY 288 -----ProHisLysGlyVal 292
DB 655 ACGGGCGGTTCGCGGACAAAGACCGCGTCGATCGCGCGGACCCCATGAAGTCGCT 714
QY 293 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyProTyrAsp 312
```



```
Db 115 CTCACCC---AATCTCTACGTACTAGCGCGACTACCTCGTCCATGGCG 768
Qy 313 GlySerGlyAspValTrpAlaPheSerValThrSerGlyThrTrpThrArgIle--- 331
Db 769 GTACAGTTTCGGAGAGTTCGGCGAGAACGCGACCTCGCGCGCTTGGGACGACATTACT 828
Qy 332 -----SerProValProSerThrAspThrAlaAsnAspTrp:PheGly 345
Db 829 CCCCGCTCGGCACTCTGCTGCTGCCCGGTACAAACACGAGCGTTCCCTGCGGGCGGA 888
Qy 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365
Db 889 TTTTGGGTCTCAGCGTCGACGCGACCAACCCACCGCTCTCGTCTCATCAC---CTC 945
Qy 366 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyAlaThr 382
Db 946 GACCGCGACCCGCGCGCCCTCGACAGCATCTACCTCTCAACCGATGCCGCGCGGACC 1005
Qy 393 TrpThrArgIleTrpAspTrpSer-----TyrPro 393
Db 1006 TGGAGGAGGTACCCAGCTCTGTCCTCCCGTCCAACTCGAAGTAACTGGGGCCACCG 1065
Qy 394 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 413
Db 1066 ACTAACGCGCGCGGTAC---AAGGACGCGACGCTGTTCGTGGCTCGACTTCAACAAC 1122
Qy 414 GlnPro-----AsnProValProSerPro-----LysLeuGly 425
Db 1123 GGTCCCAAGTGGGGGGATAGCGTGGCGCGACGCTAGCGCGGCTCACCAGTTTGGC 1182
Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTrpGlyThr 445
Db 1183 TGGTGGATGAGCGCTGTCTATCGATCCGTTCACCCCGAGCACCTGATGTACGGCAG 1242
Qy 446 GlyAlaThrLeuTrpAlaThrAsnAspLeuThrLys-----TrpAspSerGly 461
Db 1243 GGGGCGACCATCTGGCGCGCGACGACGCTCTCCCGTGTGAGAGGACTGG----- 1293
Qy 462 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla 477
Db 1294 -----GCGCGAGCTGTGTACCTCCAGATCGACGGTATCGAGGAGAAATCG 1338
Qy 478 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497
Db 1339 ATCTGTCTCGCTCCGCTCCCAAGAGCGCGCGGCTCTCTGTGCGGATCGGTGACATT 1398
Qy 498 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 517
Db 1399 AGCGGATGAGACGACGACGACTCACCAAG---CCCCAGAGATGTTTGGTGGCGCCAG 1455
Qy 518 PheThrThrGlyThrSerValAspTrpAlaGluLeuAsnProSerIleValArgAla 537
Db 1456 TTCTCCAACTCGACAGCATGACGCTCGGGCAACTTCCCAAGTGTGTGTCGCGGCC 1515
Qy 538 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 557
Db 1516 GGATCTCTCGGGACACGAGTACGACAGCGCGTGGCGCGGCTGTGCTACGCGACTGACGGC 1575
Qy 558 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThr----- 572
Db 1576 GGAGACGGGTGACCATCTTCTTACCTGCGCTCTGCGCATGACGCGCGCACTACACAG 1635
Qy 573 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 590
Db 1636 GGCACACGATTGTCAGTCGACGCGGAGCGGACGACGATCGTGTGTGTCACCAAGCTTAC 1695
Qy 591 AspProGlyGlnProValValTrpAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
Db 1696 GAGCAGGCGCTCGGAGCGGTGCTACTCGCACGACTATGGCAAGACGCTGTCT----- 1746
Qy 611 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 624
```

```
1747 ---GTTCCCGCTGGCGACCTGAAGCCCGAGACTGCCAATGTGCTCTCGGACAAAGTCCAG 1803
Qy 625 ProLysThrPheTrpAlaLeuSerAsnGlyThrPheTrpArgSerThrAspGlyGlyVal 644
Db 1804 GATGGCAGTTTCTACGCTACCGATCGCGCGCAAGTTCTTCGCTCTGACCGACGCGGGAAG 1863
Qy 645 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 664
Db 1864 TCGTATGCGCGCAAGGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 1917
Qy 665 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTrpHis 683
Db 1918 GTGAACCCCTCGGTGGCGCGGACGCTCTGGGTGCTGTTCGCGAGGGCGGTCTCTTCCAC 1977
Qy 684 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702
Db 1978 TCGACCGACTTTGGCGCTCTGTTACGAGGGTGTGTACCGCAACGCGACCTCTGTGAGC 2037
Qy 703 ValGlyPheGlyLysSer-----AlaProGlySerSerTrpProAlaValPhe 718
Db 2038 GTCGCGCGCCCAAGTCCAGTTCGAGCGCAAGAGGCTAGCGCGCTCCGCGGTCTTC 2097
Qy 719 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 736
Db 2098 ATCTGGGACCGCACAGCTGGAAGCAGCATCGGCTGTACCGCTCCGACGACACGCGC 2157
Qy 737 ThrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756
Db 2158 AGCACTCGACGCGCTCAATGACGAGGACCACTACTCTCGGCG---CCCACCATGATC 2214
Qy 757 ThrGlyAspHisAlaAsnLeuArgValTrpIleGlyThrAsnGlyArgGlyIleVal 776
Db 2215 GAGCGCGACCCCAAGTCTACGGCGCTGTATCTAGGACGACGCGCGGTATCTCGT 2274
Qy 777 TyrGlyAspIle 780
Db 2275 TAGCGCGACTT 2286

RESULT 12
ADR90295
ID ADR90295 standard; DNA; 2755 BP.
XX
AC ADR90295;
DT 16-DEC-2004 (first entry)
DE
KW endo-type xyloglucanendohydrolase;
KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
KW plant cell differentiation; precursor; ds; gene.
XX
Geotrichum sp. M128 xyloglucanendohydrolase precursor DNA.
endo-type xyloglucanendohydrolase;
xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
plant cell differentiation; precursor; ds; gene.
Geotrichum sp. M128.
Key Location/Qualifiers
CDS 202..2532
/*tag= b
/product= "Geotrichum sp. M128 xyloglucanendohydrolase
precursor protein"
202..261
/*tag= a
/label= Signal_peptide
262..2529
/*tag= c
/product= "Geotrichum sp. M128 xyloglucanendohydrolase
mature protein"
sig_peptide
mat_peptide
JP2004261037-A.
24-SEP-2004.
28-FEB-2003; 2003JP-00053286.
XX
```



```
QY 517 ValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArg 536
Db 1693 ACCTTCAGCAGCAGCGCGCTGACCTTCGCGGGCGCGGCAATGCTCGCGCGC 1752
QY 537 AlaGlySerPheAsp-----ProSerSerGlnProAsnAspArgHisValAlaPheSer 554
Db 1753 GTCGGCGCAGCAGCACCCTCTGACCGCGCGCTCGCGGGGTGACGACGAGCGCGGTACACG 1812
QY 555 ThrAspGlyGlyLysAsnTrp-----PheGlnGlySerGluProGly---GlyValThr 571
Db 1813 ACCAACACGCGCGCAGCAGCTGACGCTCTTTTCAGACCTGTGTCCCGCAGCTCGAGTCCGT 1872
QY 572 ThrGlyGlyThrValAlaAsaSerAlaAspGlySerArgPheValTrpAlaProGlyAsp 591
Db 1873 AACGCGCGCAGCAGTCCGCGTCCGCGCGCAGCGCAGCTTTGTGTGTCGCCACGACG 1932
QY 592 ---ProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610
Db 1933 GCGGACGCGCAGGGCGGTACACATCGAGCGACTACGCGAAGACGTGGACTGCGCGCAGC 1992
QY 611 GlyValProAlaAsnAlaGln---IleArgSerAspArgValAsnProLysThrPheTyr 629
Db 1993 GGCCTGAGCAAGCAGACGACAGCTATCGCGCGCAGCCGCGTGCAGCGCCACACATCTTAC 2052
QY 630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649
Db 2053 GTGTACGTGAGGGCGCATCTTCTGCTCGACCGATCGCGCAGAGCTACACCAAGAG 2112
QY 650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly--- 668
Db 2113 GGCAACGCGCTCCGCTGCTGTGAGC-----TACACGGGCGACCCCGTCAAG 2160
QY 669 -----LysGluGlyAspLeuTrpLeuAlaSerSer---GlyLeuTyrHisSer 684
Db 2161 AGCAACTTGCCTGCGCGCGAGCTGTGGTCTCCGTCAGGGCGTGTGGCATCTACCAAGC 2220
QY 685 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 704
Db 2221 ACCGACTTTGGCAACACATTCACCGCGCTTGGGGCTCGGGCTCAGTCTCAAT----- 2274
QY 705 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThr----- 722
Db 2275 -----CCCGCTGTGTTCAGCATCGGCGCGCGCAG 2304
QY 723 -----IleGlyGlyVal-----Thr 727
Db 2305 ACGCCAACGCCAGCAGCGCTCTCTCGGGCATCCGCTCGGCGTCCGAGCCCGAG 2364
QY 728 GlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspGlnHis 747
Db 2365 GGCCTGTATATGTCGACCGACAACCGCGGGTGTGGACGCGCCCTCAACGACGCGCGCAG 2424
QY 748 GlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyr 767
Db 2425 AACTATGTGGC---GCCACCGTGTATCAGCGCGGATCCCGCATCTACGCGCGGTCTAC 2481
QY 768 IleGlyThrAsnGlyArgGlyIleVal 776
Db 2482 ATTGGCATGACGCGCGCGGCATCATC 2508
RESULT 13
ID ADR90297
ID ADR90297 standard; DNA; 2268 BP.
XX AC ADR90297;
XX AC ADR90297;
DT 16-DEC-2004 (first entry)
DE Geotrichum sp. M128 xyloglucanendohydrolase mature DNA.
XX endo-type xyloglucanendohydrolase;
KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
```

```
KW plant cell differentiation; mature; ds; gene.
XX Geotrichum sp. M128.
OS
XX Key Location/Qualifiers
FH CDS 1..2268
FT /*tag= a
FT /partial
FT /product= "Geotrichum sp. M128 xyloglucanendohydrolase
FT mature protein"
FT /note= "No start or stop codon"
XX JP2004261037-A.
XX 24-SEP-2004.
XX 28-FEB-2003; 2003JP-00053286.
XX 28-FEB-2003; 2003JP-00053286.
XX (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.
XX WPI; 2004-665466/65.
XX P-PSDB; ADR90298.
XX Novel xyloglucan hydrolase derived from Geotrichum genus, having ability
XX to hydrolyze beta-1,4-glycoside linkage of glucose residue in
XX xyloglucan, useful for elucidating structure or function of xyloglucan.
XX Claim 7; SEQ ID NO 14; 86pp; Japanese.
XX The invention relates to a novel xyloglucanendohydrolase (xyloglucan-
XX specific endo-b-1,4-glucanase) enzyme derived from the Geotrichum genus
XX and having the ability to hydrolyse the beta-1,4-glycosidic linkage of a
XX glucose residue, but not that of a xylose residue contained in a
XX xyloglucan. The polypeptide of the invention may be useful for
XX elucidating the structure or function of a xyloglucan having an important
XX role in the differentiation of a plant cell. The current sequence is that
XX of the Geotrichum sp. M128 xyloglucanendohydrolase mature DNA of the
XX invention.
SQ Sequence 2268 BP; 413 A; 795 C; 717 G; 343 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2,666-26 Length: 2268
Score: 979.00 Matches: 264
Percent Similarity: 48.43% Conservative: 122
Best Local Similarity: 33.12% Mismatches: 279
Query Match: 19.07% Indels: 132
DB: 13 Gaps: 33
US-09-917-376-1 (1-957) x ADR90297 (1-2268)
QY 57 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76
Db 22 GTGACGATCTCGGTTGGCGGCTTCATCTCGGGCCTCGTGGCACACCCGACCGAAGGAC 81
QY 77 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaValArg 96
Db 82 CTGATCTACGACGCGACACATTCGCGGCGACGTACAGGTGGAAACGCGCGCAAGTGGAG 141
QY 97 TrpIleProLeuLeuAspTrpVal-----GlyTrpAsnAsnTrpGlyTyr 111
Db 142 TGGGAGCCCATCACCGACTTTTATCATCAACAACGCGCTCGGGGCAACGCGCGCAACCTG 201
QY 112 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131
Db 202 CTCGGCAGGAGAGCATTCGCTCGACCCCAACACCGACGCGCTGTACCTCGCGCAA 261
QY 132 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 151
Db 262 GCGCACTATGTG---CAGTGGGACCGC---TGGGCGGCGCTTCTCTCTCGGACGACGCG 315
```


152 GlyAlaThrTrpGlnIleThrProLeuProPheLeuGlyValMetProGlyArg 171
172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191
376 AATGCGCGGAGCGCTCGCGGTCAACCGCACTGGACGAGCTGTGGTTGGCTCG 435
192 ProSerGlyLeuGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211
436 CGCAAG--CAGGGCTGTGGCGCTCCACTGACCGCGCTCAGACGTGTGGCGATGAAC 492
212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 231
493 CAGCTCCCGCAGAGCTCAGCTGATC-----GGATC----- 522
232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 251
523 -----GGATCATCTCGGTCTATCTCGACCCCAAGACGTCGGCAGCATACGTC 573
252 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheThrSerArgAsp 271
574 GCTTCGACCGCTCGCGGGGTGTGG-----GTCACTGG-----GAC 612
272 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThr----- 284
613 GCGCGGCGCAACTGTGTCCAGGTGGCGCGCAGCTTACCGACGTGTGACAGTGGACAAG 672
285 -----GlyPheIleProHisLysGly 291
673 AGCATGTGTCGTGCGCGCACCGGATCCAGAGCTCGGGGCCCTCCCGATCAAGATC 732
292 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 311
733 GCACGTGGCAAG--AACGGCAGGCTGTATACCTACTCGACGCGACCGGCGCCCTGG 789
312 AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle 331
790 GCGGTCTCTACGGCGAGGTCTGTCTACGATCCACCAACCGCACTGSAAGCACAATC 849
332 SerProValProSerThrAspThrAlaAsnAspTyr----- 343
850 ACC-----CCCTCGCGCGGCGGCGCAACAGTACCTCGCCCGCACTGGCAACAAAAG 903
344 -----PheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360
904 GTCGTTCGCGAGGCTGGAACGATCTCGGTGCT-----AACGGCAGACACGTCGTC 957
361 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 380
958 GTGTCAGCGCTCGACGCAACCGCGAAGACTCA--GTGTACTCTCTCGCGCGACGCGCGC 1014
381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400
1015 AACTGTGGAAGACCTCGGCAAGCTGACACGCGCGCGCGCGCGCGCGCACTCGCAG 1074
401 LeuAspIleSerAlaGlu-----ProTrpLeuThrPheGlyVal 413
1075 AAGAGTTCGCGCAAGCTGCGCAACGCGCATCCCGCTCCCTGGCTCTCGTTCAGAAC 1134
414 GlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433
1135 CGCGGCTCGGTATTGTC-----GGATTGCGGTGTGGCTCGCGCGCATCTCTCTC 1185
434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453
1186 GACCCCTTC--TCGACCGCTCTGTACGCGCACAGGCGCTGTGATCTGGCGCACTGAC 1242
454 AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMet----- 469
1243 CCGGTGTGGCGCGCGACTCG-----AACGAGCGCGCGAGCTGGTACATCAAC 1290
470 ValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGly----Ala 488

1291 ACAGAGGCGCATCGAGGAGACGCGCATCTGTGTCCTCAAGTCGCGCGCTCGGGCGCGCG 1350
489 ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal 508
1351 CACCTCTTCAGCGGATGTACGACCTTCGGCGGCATCGCGCATGACTTC---TCGCTC 1407
509 ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGlu 528
1408 CCGAGCCCATGTACAGCAAGCGACCTTCAGCAGCAGCGCGCTCGACTTTGCGGC 1467
529 LeuAsnProSerIleIleValArgAlaGlySerPheAsp-----ProSerSerGlnPro 546
1468 CCGCGCGCAATGTGTCTCGCGCGCTCGCGCGCAACGACCATCTCTGACGCGCGCTCGC 1527
547 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp-----PheGlnGly 564
1528 GGTGTCAGCAGCGCGCGTACAGCAACAGCGCGCAGCTGGAGCTCTTTTCAGACC 1587
565 SerGluProGly--GlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer 583
1588 TGTGTCCCGAGCTCGAGTGGTACGCGGCGCACGATCGCGTCGCGCGCGGCAAG 1647
584 ArgPheValTrpAlaProGlyAsp---ProGlyGlnProValValTyrAlaValGlyPhe 602
1648 ACCTTTGTGTGTGTCGCCCAAGCGCGCAAGCGCGCGTACACATCGAGCGACTAC 1707
603 GlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln---IleArgSerAsp 621
1708 GGCAGACGTGACTCGCGCGAGCGCTGAGCAAGCAGCAGACAGATATCGCGCGCGAC 1767
622 ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp 641
1768 CCGGTGACGCCAACACATCTCTACGTAGTACGAGCGGCGACTTCTTCGTCTCAGCCAT 1827
642 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 661
1828 GCGCGCAAGAGCTACACCAAGAGGCGCAACGCGCTCCCGTCTGCTGCGACG- 1878
662 MetPheHisAlaValProGly-----LysGluGlyAspLeuTrpLeuAlaAla 677
1879 ---TACACGCGGACGCGCGCTCAGCAGCACTTGTGTCGCGCGAGCTGTGGGTCTCGCTC 1935
678 SerSer--GlyLeuTyrHisSerThrAsnGlySerSerTrpSerAlaIleThrGly 696
1936 AAGCGGTGTGGCATCTACACAGCAGCTTTGGCAACACATTCACCGCGCTTCGCGGC 1995
697 ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAla 716
1996 TCGGGCTCCAGTCTCAAT-----CCGCT 2019
717 ValPheValValGlyThr-----IleGlyGly 725
2020 GTGTTGAGCATCGCGCGCGCAGACGCCCAACGCCAGCAGCGCTCTCTCTCGGGGC 2079
726 Val-----ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrp 739
2080 ATCCCGTTCGCGCTCGCAGCCCGAGGCGCTGTATGTGCGACCAACCGCGGTGTGG 2139
740 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 759
2140 ACGGCGCTCAACAGCAGCGCGCACTATGTGGTGGC---GCCACCGTGTATCAGCGCGCAT 2196
760 HisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776
2197 CCGCGCATCTACGCGCGCTCTACATTGGCATGAACGCGCGGGGCATCATC 2247
RESULT 14
ADR90301
ID ADR90301 standard; DNA; 2274 BP.
XX
AC ADR90301;
XX

QY	454	AspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMet-----	469
DB	1246	GCCGTGTCGCGCGGACTCG-----	1293
QY	470	ValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGly---Ala	488
DB	1294	ACAGAGGATCGAGAGCGGATCTGTCTCAAGTCGCGCCTCGCGGCCCGCG	1353
QY	489	ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal	508
DB	1354	CACCTCTTCAGCGCATCTAGACCTCGCGCGCATCGCCATGATGACTTC---TCCGTC	1410
QY	509	ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGlu	528
DB	1411	CGCGAGCCCATGTACAGAACGCCACCTTCAGCAGCAGCGCGCTCGACTTCGGGCG	1470
QY	529	LeuAsnProSerIleIleValArgAlaGlySerPheAsp-----ProSerSerGlnPro	546
DB	1471	CGCGGCGCAATGTCTCGCGCGCTCGGCGCGCACGACCATCTCTGACGCGGCGTCGCG	1530
QY	547	AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp-----PheGlnGly	564
DB	1531	GGGTGACGACGCGCGGTACAGCACCAACAGCGCGCACAGCTGCGCTCTTCAGACC	1590
QY	565	SerGluProGly---GlyValThrThrGlyThrValAlaAlaSerAlaAspGlySer	583
DB	1591	TGTGTCCCGACGCTCGGTAAACGCGCGCACGATCGCGTTCGCGCGCGCAAG	1650
QY	584	ArgPheValTrpAlaProGlyAsp---ProGlyGlnProValValTyrAlaValGlyPhe	602
DB	1651	ACGTTGTGTGTGTCGCCCGCAGAGCGGCGGCAAGGCGGCGTACATCTGAGCGACTAC	1710
QY	603	GlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln---IleArgSerAsp	621
DB	1711	GCGACAGCTGTGACTGCGCGGAGCGGCTGAGCAGCAGCAGGATATCGCGCGCGAC	1770
QY	622	ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp	641
DB	1771	CGCGTGCAGGCCAACACATCTTACGTGTACGTGCGGCGGCGAGCTGTGGTCTCGCT	1830
QY	642	GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal	661
DB	1831	GGCGGCAAGAGTACACCAAGAGGCAAGCGGCTCCCGTCTGCTGCGACG	1881
QY	662	MetPheHisAlaValProGly-----LysGluGlyAspLeuTrpLeuAlaAla	677
DB	1882	---TACCGGCGCGCGCTCAGCAGCAACTTGGTGGCGGCGAGCTGTGGTCTCGCT	1938
QY	678	SerSer---GlyLeuTyrHisThrAsnGlySerSerTrpSerAlaIleThrGly	696
DB	1939	AAGGGGTTGGCATCTTACACAGCAGCGACTTTGGCAACACATTCACCGCTTCGGGCG	1998
QY	697	ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAla	716
DB	1999	TCCGCTCCAGTCTCAAT-----CCGCT	2022
QY	717	ValPheValValGlyThr-----IleGlyGly	725
DB	2023	GTGTTTCAGCATCGCGCGCGCGCGGAGCCCAACCGCCAGCAGCTCTTCTCTCGGCG	2082
QY	726	Val-----ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrp	739
DB	2083	ATCCCGTCGCGTCGCGAGCGCGGCGCTGTATATCTCGACCGACACGCGGGTGTGG	2142
QY	740	ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp	759
DB	2143	ACGCGCTCAACGACGCGCACCAATATGTGTGGC---GCCACCGTGTATCAGCGCGAT	2199
QY	760	HisAlaAsnLeuArgValTyrIleGlyThrAsnGlyValArgGlyIleVal	776
DB	2200	CCCCGATCTTACGCGCGCGTCTTACATTTGGCATGACGCGCGCGGCGATCATC	2250

RESULT 15	
ADD42054	
ID	ADD42054 standard; DNA; 5698 BP.
XX	
AC	ADD42054;
XX	
15-JAN-2004	(first entry)
XX	
Trichoderma reesei Family 74 xyloglucanase-encoding DNA, SEQ ID NO:1 #1.	
XX	
Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;	
KW	xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;
KW	ethanol production; detergent composition; fabric treatment;
KW	textile treatment; ds.
XX	
OS	Hypocrea jecorina.
XX	
PN	W02003089598-A2.
XX	
30-OCT-2003.	
XX	
17-APR-2003; 2003WO-US011831.	
PF	
19-APR-2002; 2002US-0373987P.	
PR	
XX	(NOVO) NOVOZYMES BIOTECH INC.
PA	
XX	Michael R, Zaretsky E, Haas J;
PI	
XX	
WPI; 2003-845528/78.	
DR	P-PSDB; ADD42055.
XX	
New polypeptides having Family 74 xyloglucanase activity, and encoding	
PT	nucleic acid molecules, useful for degrading cellulose- and hemicellulose
PT	-containing biomass to ethanol or as a detergent.
XX	
Claim 1; SEQ ID NO 1; 96pp; English.	
XX	
The invention relates to a Family 74 xyloglucanase (ADD42055) from the	
CC	fungus Trichoderma reesei (Hypocrea jecorina), and nucleic acids encoding
CC	it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic
CC	linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.
CC	The invention also relates polypeptide sequences at least 70% identical
CC	to the enzyme, expression vectors and host cells comprising a nucleic
CC	acid of the invention, the recombinant production of the enzyme, and
CC	mutant enzymes and the nucleic acids encoding them. The xyloglucanase of
CC	the invention can be used in the degradation of cellulose- and
CC	hemicellulose-containing biomass to produce ethanol. It can also be used
CC	in a detergent composition for treating fabric during a machine washing
CC	cycle. The present sequence represents a Trichoderma reesei Family 74
CC	xyloglucanase-encoding DNA. Note: the present sequence (referred to as
CC	SEQ ID NO:1) is given as shown in the sequence listing. A comparison of
CC	this sequence with that shown in figure 1 (also referred to as SEQ ID
CC	NO:1) shows that the present sequence is twice as long as the figure 1
CC	sequence (ADD42060) and contains alternate 60 bp blocks of sense strand
CC	and complementary strand sequence.
XX	
Sequence 5698 BP; 1145 A; 1704 C; 1704 G; 1145 T; 0 U; 0 Other;	
Alignment Scores:	
Pred. No.:	1.36e-24
Score:	940.00
Percent Similarity:	30.11%
Best Local Similarity:	22.08%
Query Match:	18.31%
DB:	10
US-09-917-376-1 (1-957) x ADD42054 (1-5698)	
QY	25 SerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
DB	128 TCTCGATCTCTTGGCTTGTCTTGGCGGCGCTC-----ATCCCTGCC 169

QY 45 HisAlaAlaThrThrGln----- 50
|||||
Db 170 CATGCTGCTTCTTCCAGAGAGCTCAGGAACGGGAACAGGACCCCGCGCAGTAGGACGG 229
QY 51 -----ProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGly 67
|||||
Db 230 GTACGACGGGAATTCATGGAAGAACGTCAAGCTCGCGCGCGCGGCTTCTGTCGCCGCGC 289
QY 68 IleValPheAsn-----GluGlyAlaPro----- 75
|||||
Db 290 ATCATCTTCCAAAGTACCTTCTTCAGATTTCGAGCGCGCGCGCGCGCGCGGACG 349
QY 76 -----GlyIleLeuTyrValArgThrAspIleGlyGlyMet 87
|||||
Db 350 TAGTAGAAGGTTCCCAAGAACAAAGGCGTGCATATGCACGAACAGATTTGGCGGGCTG 409
QY 88 TyrArgTrp----- 90
|||||
Db 410 TACCGCTCAAAAGGTTCTGTTTCCGCATCGTATACGTGTTGTCATATACCGCGCGAC 469
QY 91 -----AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal----- 104
|||||
Db 470 ATGGCGGAGTTTCGCGCAGACTCATGGACCGCGCTCACGGATGGGATTTGCTGATAATGCC 529
QY 105 GlyTrp-----AsnAsnTrpGly----- 110
|||||
Db 530 GGCTGGCAACAGCGGCTGCTGAGTACCTGCGCGGCGAGTGCCTAACCTAACGACTATTACGG 589
QY 111 -----TyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnVal 127
|||||
Db 590 CCGACCGTGTCTGGGGCATCGACGCTGTTGGCTTGTATCCGCAGGACGATCAAAAGGTG 649
QY 128 TrpAlaAlaVal----- 131
|||||
Db 650 TATGCGCAGTGAACCCGCTAGTCGCAACACGCGAACTAGGCGTCTGCTAGTATTTCAC 709
QY 132 -----GlyMetTyrThrAsnSerTrpAsp----- 139
|||||
Db 710 ATACGGGTCTCGGCATATATACGAACAGCTGGTC-TGTGATGTCTCTCAGATCTAGACC 768
QY 139 ----- 139
Db 769 TATGATTGGACGGCGGTACATATGCTTCTGCACCAACACTACAGAGAGTCTAGATCTGG 828
QY 140 -----ProAsnAspGlyAlaIleLeuArg 147
|||||
Db 829 ATACTAACTTCCTGACATTTGCCCATATATAGGAGTCCGAGTAATGGAGCCATCATTCGC 888
QY 148 SerSerAspGln----- 151
|||||
Db 889 TCGTCAGACCGGACTGTAAACGGGTATATATCCCTTAGGCTCATTTACCTCGGTAGTAAGCG 948
QY 152 -----GlyValaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGlyVal 167
|||||
Db 949 AGCAGTCTGGGGCGGCGCAACGTGTCTTCCAACTTGCCCTTCAAAGTCGGGGGTAAAC 1008
QY 168 MetProGlyArg----- 171
|||||
Db 1009 ATGCCAGACCGCGGTTGCACCAAGAGTGTGTTGAACGGGAAGTTTCAGCGCCCATTTG 1068
QY 172 -----GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 187
|||||
Db 1069 TACGTCCTCGGGAGCGGAGAGCGTCTGGGTGTCGATCCGACACTTCCGCGCAACTCCAAACATCATC 1128
QY 188 TyrPheGlyAlaPro----- 192
|||||
Db 1129 TACTTTGGTCTCTCGCGCTCTCCAGACCGACAGCTAGGCGCGGTTGAGGTTGTAGTAG 1188
QY 193 -----SerGlyIysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 207
|||||
Db 1189 ATGAACACCGACGCTCAGGAACCGGCTCTGGAAGTCTACGGAACGGCGGCTGACCTTT 1248
QY 208 SerGlnMet----- 210
|||||

Db 1249 TCCAAAGTCTCGGCGAGTCTTTTGGCGGAGACCTTCAGATGCTGCGCGCAGCTGGAAA 1308
QY 211 -----ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 227
|||||
Db 1309 AGGTTCCAGAGCTCGTTTCCAGCAACTGGGACGTATATCCAGACCCGAGTGATTCCAAAC 1368
QY 228 GlyTyrGln----- 230
|||||
Db 1369 GGCTCAACAGCAGCAAGTGGCGTTGACCTGTCATGTAGGGTCTGGGCTCAGTAAGGTTG 1428
QY 231 -----SerAspIleGlnGlyValValTrpValAlaPheAspIysSerSerSer 247
|||||
Db 1429 CCGATGTTGTCGACACGAGGACTCATGTGGTTACGTTTCGACTCAACACGACGACG 1488
QY 248 LeuGlyGlnAla----- 251
|||||
Db 1489 ACCGGGGAGCCCTGTTCTCTGAGTACACCAATGCAAGCTGAGTTGGTCGTCGTGC 1548
QY 252 -----SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe 267
|||||
Db 1549 TGGCCCCCTCGGACGCTCGTATCTTTGTGGACGGCTGATACATCACTGCTTCAGTC 1608
QY 268 Trp----- 268
|||||
Db 1609 TATGTGACGACGTGCAGACATAGAAACAAACCGTCCGACTATTGTAGTGACGAAGTCAG 1668
QY 269 -----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286
|||||
Db 1669 ATPACACTCGTGAATGCGGCTCCACGTGAGTGTGTACCGGGGCGACGAGGAATAC 1728
QY 287 IleProHis----- 289
|||||
Db 1729 TTTCTCTCAAGTTACGCGGAGTGCACCTCAGACATGCGCCCTCGGTCCTCTTATG 1788
QY 290 LysGlyValPhe-----AspProValAsnHisValLeuTyrIleAlaThrSerAsn 306
|||||
Db 1789 AAAGAGTGTTCGCGAACTGCAGCCAGCAGAGAGGCTTGTATCTGACCTATTCCGAT 1848
QY 307 ThrGlyGlyPro----- 310
|||||
Db 1849 GGCACAGGCGCGCTTTGACGTCGTCGCTCTTCCGGAACATAGACTGGATAAGGCTA 1908
QY 311 -----TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326
|||||
Db 1909 CGGTGTCGCGCTATGATGGCACACTTGGCTCAGTGTGAGGTACGACATTCAGCGGGGA 1968
QY 327 ThrTrpThrArgIleSerProValProSerThrAspThr----- 339
|||||
Db 1969 ACTTGAAGACATACTACCGTGTGNAACGAGTCAACCTCCATGCTGTAAACGTCGCCCT 2028
QY 340 -----AlaAsnAspTyrPheGlyTyrSerGlyLeu 349
|||||
Db 2029 TGAACCTTTCTGATCACCCCTGTCCTCGGATCAGATCTATACTTTGGCTTTGGCGGCTT 2088
QY 350 ThrIleAsp----- 352
|||||
Db 2089 GGCCTCGATTGTAGTGGGACAGAGACCTAGTCTAGATATGAAACCGAAACCGCCGGAA 2148
QY 353 -----ArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpPro 369
|||||
Db 2149 CCGGAGCTAAACAAAGCCAGAACCTTGTGTGTGCTCTTTTGAACCTTTGTGTGCCA 2208
QY 370 AspThr----- 371
|||||
Db 2209 GATGTCACGCTGTTTTCGTCCTTGGGAACAACAACGAGAACTTGAAGAACCCCGGT 2268
QY 372 -----IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrp 389
|||||
Db 2269 CTACGAGTCGACTTTTCGTCGACCGACTCTCGGACAACTAGGACCGGATCTGGCGCTGG 2328
QY 390 ThrSerTyrProAsnArgSer---LeuArgTyrValLeuAspIleSerAlaGluPro--- 407
|||||

Db 2329 GCGAGCTATCGAAAGCCAGCTGGCTGAGACC-CTGTTGTACCTCGGCTAGACCCGCGAC 2387
Qy 407 ----- 407
Db 2388 CCGCTCGATAGGCACTGAGACCTATTACTACAGCATCTCAGTGAGTCACCTTTAAACGATC 2447
Qy 408 -----TrpLeuThrPhe-----GlyValGlnProAsn----- 416
Db 2448 CGATGCGAATGGCTGACTCTGGATAATGATCTCGTAGAGTCACCTCAGTGAGAAATGCTAG 2507
Qy 417 -----ProProValPro----- 420
Db 2508 GCTAGCTTACCGACTGACTCTGCTTCAAGACTCCCAAAGCACCGTGGATCAAGAACAC 2567
Qy 420 ----- 420
Db 2568 TTTATCGATGTGATGACTGAGACGAAGTTCTGAGGGTTTCGTGGCACCTAGTCTTGTGTTG 2627
Qy 421 -----SerPro-----LysLeuGlyTrpMet 427
Db 2628 AAATAGCTTACACTCGAGCGAGTCAACCGTCCGATGCTCATCAAGCGCTCGGCTGGATG 2687
Qy 428 AspGluAlaMetAla----- 432
Db 2688 ATTGAGTCTCTCGGCTCGCTCAGTGGCAGGCTACCAAGTAGTTCCGCGGACCGACCTTAC 2747
Qy 433 -----IleAspProPheAsnSerAspArgMetLeuTyThrGlyAla 447
Db 2748 TAACTCAGAGAGCAGATTGACCCACCGACGACCAACCACTGGCTCTACGGCACCGGAATG 2807
Qy 448 ThrLeuTyAlaThr----- 452
Db 2808 ACAATCTTTGGCGTCTAACTGGTTGGCTGTGTTGGTGACCGAGATCCGCTGGCTTAC 2867
Qy 453 -----AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisAla 467
Db 2868 TGTAGAAACCGCGCCAGATCTCAACACTGGGACACCGCCACAAATGTGTCAATCCAA 2927
Qy 468 ProMetValLysGly----- 472
Db 2928 TCATGGCAGACGCGGTGCTAGAGTGGTTGACCTGTGCGCGGTGTACACAGTTAGTT 2987
Qy 473 -----LeuGluGluThrAlaValAlaAsnAspLeuIleSerProSerGly 487
Db 2988 AGTGACCGCTGCGCATCGAGGAATCTCCGTCAGAGACCTGGCTCTGACCCCGCGGA 3047
Qy 488 AlaProLeuIleSer----- 492
Db 3048 AGCGAGCTATTGGCGTAGCTCCTTAAGAGGCAAGTCTTGACCGGAGACGTTGGCGCGCT 3107
Qy 493 -----AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAla 507
Db 3108 TCGCTCGATAACCCCGCAGTCGAGACGACCAACGCGTTCCACCTTGGCCAGCAACGAC 3167
Qy 508 Val----- 508
Db 3168 CTCGGGACATCGCGCGCTCAGCCTCTGCTGTTGCCGAAGTGAAACGGTCGCTTTGCTG 3227
Qy 509 -----ProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524
Db 3228 GAGCCCTGTAGCGCGACGAGCTCTGGGCAACGCCACATGGGCCACCTCGACGAGCGTC 3287
Qy 525 AspTyAla----- 527
Db 3288 GACTACGCGCGGAGGCTCTGCCAGACCCGTTGCGGGGTGTACCCGGTGGAGCTGCTCGCAG 3347
Qy 528 -----GluLeuAsnProSerIleLeValArgAlaGlySer----- 539
Db 3348 CTGATGCGGCCCTACTCGGTCAAGAGCGTCGTCGCGTCCGCAACACACCCCGCGCAGCAA 3407
Qy 540 -----PheAspProSerSerGlnProAsnAspArgHisValAla 552
Db 3408 CAGGTGGCCATCTTGAGCCAGTCTTCGACGACGCGCGCTGTGTGGAGCCGTCCTGTCGCGCGCTGTCGTT 3467

Qy 553 Phe-----SerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 569
Db 3468 GTCCACCCGTAGACGCTCCGACGCGCGCGAGCTGGAGCATCGACTACGCGCGCACACG 3527
Qy 570 ValThrThrGlyGly----- 574
Db 3528 TCCATGAACGGCGCAGGCTGCCGCGCTGCACCTCGTAGCTGATGCCGGCTGTGC 3587
Qy 575 -----ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589
Db 3588 AGTACTTTCGCGCAGCGTGGCTTATTCGCGCGACGCGACACGATCTCTGTCGACC 3647
Qy 590 GlyAspProGly----- 593
Db 3648 GCCTCGTCGGGCGGTGCCACCGGATAAGCCGGCTGCCGCTGTGTAGGAGACCACTGG 3707
Qy 594 -----GlnProValValTyAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGly 611
Db 3708 CGGACGAGCGCTCGCAGCGCTCG-----CAGTTCCAGGCGAGCTTTCCTCGTCTCGAGC 3764
Qy 612 ValProAlaAsnAlaGln----- 617
Db 3765 CTGCCCGCGGCGCGCAGCTCGCGAGCTCAAGSTCCCTCGAAACGCGAGCAGAGCTCG 3824
Qy 618 -----IleArgSerAspArgValAsnProLysThrPheTyAlaLeu 631
Db 3825 GACGGGCGCCCGGCTCATCGCTCGGACAGAACACACAGCGTCTTCTACGCGCGC 3884
Qy 632 SerAsnGlyThrPhe----- 636
Db 3885 TCCGATCGACCTTTTAGTAGCGGACCTGTTCTTCTGTTGTCGCAAGATGCGGCGC 3944
Qy 637 -----TyArgSerThrAspGlyGlyValThrPheGlnProValAlaAla 651
Db 3945 AGGCCTAGCTGGAATAACGTCAGCAAGGACACCGCAGCAGCTTC-----ACGCGC 3995
Qy 652 GlyLeuProSerSerGlyAlaValGlyValMet----- 662
Db 3996 GGG-----CCAAAGCTGGGCGCAGGAGTGCAGTGTCTCTGTGGCGCTCGTGAAGTGC 4052
Qy 663 -----PheHisAlaValProGlyLys 669
Db 4053 GCGCGCGGTTGACCCGCTCGCTCCCTCGATCGGGATATCGCTGCTACCCGACCCACC 4112
Qy 670 GluGlyAspLeuTrpLeuAlaAlaSerSer----- 679
Db 4113 GCGGGCAGCTTGTATGTCTCGACCGGCTAGGCCCTATAGCGACGAGTGGGTGTGG 4172
Qy 680 -----GlyLeuTyThrHisSerThrAsnGlyGlySer 689
Db 4173 CGCCCGTGCAACATACAGAGCTGGCTCTCGCATATTCGCTCCACAGACTCGGGCAGC 4232
Qy 690 Ser----- 690
Db 4233 ACCTTTGGCCAGTCTCCACCGCTGTAAGCCGCTATAGCGAGGTGTCTGAGCCGCTGC 4292
Qy 691 -----TrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 704
Db 4293 TGGAAACCGTTTCAGAGGTGGCGGACTCCAAACACCTACCAGATCGCCCTCGGTGTGGC 4352
Qy 705 PheGlyLysSer-----AlaPro 710
Db 4353 TCAGGCTCGAACTGGAACCTGTATGCTGTGTGGATGTCTAGCGGAGCCCAACACCCG 4412
Qy 711 GlySerSerTyProAlaValPheValValGlyThrIleGlyValThrGlyAla----- 729
Db 4413 AGTCCGAGCTTGACTTGGACATACGAGTCCGGAC-----GGCCGCTCAGGGGCTCGC 4466
Qy 730 ---TyArgSerAspAspCysGlyThrThrTrpVal----- 740
Db 4467 CTCTACGCCAGTGGAGACGACGCGGCTCTCTGGAGCCGTCGCGGCGAGTCCCGGACG 4526


```

QY 741 -----LeuIleAsnAspAspGlnHisGlnTyrGly--- 750
Db 4527 GAGATGCGGTACCTCTGTGCGCGGAGGAGCTCCGACATCCAGGGCTCCAGGGCTTC 4586
QY 750 ----- 750
Db 4587 GGCTCCATCGACAGCACCAAGGTCCGCGCAGCGCTGTAGGTCCCGAGGGTCCCGAAG 4646
QY 751 -----AsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 766
Db 4647 CCGAGGTAGCTGTCTGTGTTCAGCGCGCTCCGCGCAGCACCGCCGGG-----CAAGTC 4700
QY 767 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp----- 779
Db 4701 TACGTGGGCACCAACGCGCGGGCGCTCTTTACGCTCAGGCGTCTGGCGGCCCGTTCAG 4760
QY 780 -----IleGlyGlyAlaPro 784
Db 4761 ATGCACCGGTGTGTCGCGCCCGCGAGAAATGCGAGTCCGAACCGTCGCGCGCGGC--- 4817
QY 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804
Db 4818 ACGGGCGGGACTTCTCTCGACCAAGCAGCAGCAGCAGTACTTGGCAGCCCGCCCG 4877
QY 805 SerPro-----SerSerSerProSerProSerProSerPro 816
Db 4878 TGCCCGCCCTGAAGGAGCAGCTGGTTCGTCTCGTCTGTCTCATCTCTTCCGCCAGCTCG 4937
QY 817 SerSerSerProSerSerSerPro-SerProSerPro----- 828
Db 4938 AGCACACGCTGAGGTGAGCGTGTGTATCCACGACCGCGGCTTGGAGAAGGCGGTGAGC 4997
QY 829 -----SerProSerProSer----- 833
Db 4998 TCGTGTGCGACTCCAGCTCGCAACATAGTGTCTGGGCCCGAACGACGCTGACTTCGTGCG 5057
QY 834 ----ProSerArgSerProSerProSerAlaSerProSerProSerProSerProSerPr 852
Db 5058 AGGACCAAGCTCGGCCCGCGTCCACCGGGGTGAGGGGTGCGCGGCTGCCACTGAAGCAGC 5117
QY 852 oSerSerSerProSerSerProSerProThrProSer 865
Db 5118 TCCTGGTCGAGCGCGCGCCAGGGTGCCCACTGCCACAGC 5157

```

Search completed: October 5, 2005, 09:19:17
Job time : 1592.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 5, 2005, 07:43:21 ; Search time 9202.32 Seconds
(without alignments)
3958.515 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLTWRRRLVSL.....RASFGSNPATPDATYLOX 957

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Pgapop 6.0 , Pgapext 7.0
Delop 6.0 , delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q/cn2.1/USPTO spoel/US09917376/runat_04102005_164328_26911/app query.fasta_1.2524
-DB=EST -QFMT=fastest -SURFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MAIRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CGN 1.1.12009 @runat_04102005_164328_26911 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	631.5	12.3	704	CF872104	CF872104 trico29xo
2	631.5	12.3	761	CB902332	CB902332 trico29xo
3	603.5	11.8	747	CF876916	CF876916 trico074xd
4	603.5	11.8	814	CB905388	CB905388 trico074xd
5	415.5	8.1	707	CF880713	CF880713 trico082xn
6	415.5	8.1	782	CB907625	CB907625 trico082xn
7	401.5	7.8	693	CF882065	CF882065 trico29xo
8	362	7.0	748	CE342004	CE342004 tigr-gss-
9	361.5	7.0	751	CF868882	CF868882 trico16xm

10	361.5	7.0	803	6	CB898982	CB898982 trico16xm
11	360	7.0	791	9	CNS028QT	AL186302 Tetraodon
12	358	7.0	285	7	CF706250	CF706250 CCAGF33TF
13	358	7.0	509	7	CF707273	CF707273 CCADL40TF
14	358	7.0	741	7	CF710894	CF710894 CCACL94TO
15	356.5	6.9	1066	3	CR704524	CR704524 Tetraodon
c 16	355	6.9	522	9	CE190185	CE190185 tigr-gss-
c 17	353	6.9	493	8	CC066065	CC066065 CSU-K33r
18	351	6.8	877	9	CR071250	CR071250 Forward s
c 19	350	6.8	694	8	CC119324	CC119324 NDL.70K21
c 20	347	6.8	259	8	AZ374273	AZ374273 IM0126G21
c 21	347	6.8	587	6	C79507	C79507 C79507 Mous
c 22	340.5	6.6	708	9	AG459371	AG459371 Mus muscu
23	340.5	6.6	738	7	AG468194	AG468194 Mus muscu
24	338	6.6	738	7	CF866388	CF866388 trico06xe
25	338	6.6	794	6	CB896354	CB896354 trico06xe
c 26	333	6.5	703	9	CE176982	CE176982 tigr-gss-
c 27	329.5	6.4	476	6	C79500	C79500 C79500 Mous
c 28	329	6.4	888	7	CV069323	CV069323 WPAEHUX15
29	328	6.4	562	8	AZ855343	AZ855343 2M0159A22
30	325	6.3	807	7	CN783154	CN783154 ai_L003.4
31	322	6.3	1270	9	AG448439	AG448439 Mus muscu
32	314.5	6.1	546	2	BF072864	BF072864 NCSM3H7T3
33	314	6.1	619	8	BZ192350	BZ192350 CH230-248
c 34	314	6.1	959	9	CNS022KT	AL221078 Tetraodon
35	311	6.1	726	7	CO549464	CO549464 LYEST7094
36	310	6.0	744	8	BZ277113	BZ277113 CH230-450
c 37	307	6.0	538	9	CC847893	CC847893 NDL.4416
c 38	306.5	6.0	389	5	BU639045	BU639045 mgcw011XD
c 39	306	6.0	961	9	CNS030KT	AL256070 Tetraodon
c 40	301.5	5.9	464	8	AZ928926	AZ928926 479.H1f19
c 41	298.5	5.8	440	8	AZ312209	AZ312209 IM0027L19
c 42	297	5.8	929	7	CF885920	CF885920 trico085xf
c 43	292	5.7	393	8	AZ332252	AZ332252 IM0060H12
c 44	291	5.7	761	9	CNS041RC	AL270561 Tetraodon
c 45	290	5.6	775	9	AG603507	AG603507 Mus muscu

ALIGNMENTS

RESULT 1
CF872104
LOCUS trico29xo|3.b1 T.reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cdna clone trico29xo|3, mRNA sequence.
ACCESSION CF872104
VERSION CF872104.1 GI:38126786
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
REFERENCE 1 (bases 1 to 704)
AUTHORS Diener S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,
Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,
Ward, M. and Dean, R.A.
TITLE Characterization of the protein processing and secretion pathways
in a comprehensive set of expressed sequence tags from Trichoderma
reesei

JOURNAL FEMS Microbiol. Lett. 230 (2), 275-282 (2004)

COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: LR-F1 primer:
Location/Qualifiers
1. .704
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"

FEATURES

source
Location/Qualifiers
1. .704
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"


```

Db 424 AGTAATGGAGCCATCTTCGTCGTACAGCGCGGCAACGTCGTCTTCACCAACTTG 483
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlyArgLeuAlaValAsp 180
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 484 CCCTTCAAGTCGGGGGTAAATGATCCAGAGCGCGAGCGGAGAGCGTCTGGCTGTGAT 543
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 181 ProAsnAsnAspAsnIleLeuTy-PheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 544 CGGCGCAACTCCACATCATCTTCTGGTCTGCTCAGGAACGGGCTCTGGGAAGTCT 603
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrlle 220
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 604 ACGAGCGCGGGTGCACCTTTTCCAGGCTCTGCTTCACGGCACTGGGAGCTATCATC 663
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 221 AlaAsnProThrAspThrThrGlyTyrglnSerAspIleGlnGlyValValTrpValala 240
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 664 CCAGACCGGAGTGATTCCAAAGCGCTACCAAGCGATCAGCAAGGACTCATGTGGGTAGC 723
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 241 PheAspLysSerSerSerSerLeuGlyGlnAlaSer 252
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 724 TTTCGACTC-ACCAGCAGCAGCAGCGGNGGAGCCACG 758
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

```

RESULT 3
CF876916 747 bp mRNA linear EST 31-OCT-2003
LOCUS
DEFINITION
tric074xd02.b1 T.reesei mycelial culture, Version 6 October 2003
Hypocrea jecorina cDNA clone tric074xd02, mRNA sequence.

CF876916
CF876916.1 GI:38131598
EST.
Hypocrea jecorina (anamorph: Trichoderma reesei)
Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaaceae; Hypocrea.

REFERENCE
AUTHORS
Diener,S.E., Dunn-Coleman,N., Foreman,P., Houfek,T.D.,
Teunissen,P.J.M., van Solingen,P., Dankmeyer,L., Mitchell,T.K.,
Ward,M. and Dean,R.A.
TITLE
Characterization of the protein processing and secretion pathways
in a comprehensive set of expressed sequence tags from Trichoderma
reesei

JOURNAL
COMMENT
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: In-F1 primer.

FEATURES
source
1..747
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric074xd02"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:
Pred. No.: 8.65e-27 Length: 747
Score: 603.50 Matches: 121
Percent Similarity: 65.43% Conservative: 38
Best Local Similarity: 49.79% Mismatches: 64
Indels: 20
DB: 11.75% Gaps: 6

```

US-09-917-376-1 (1-957) x CF876916 (1-747)
Qy 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 61 TCTCAGATCCTTGGCCCTTGTCTCTGGGGCCGTC-----ATCCCTGCC 102
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 45 HisAlaAlaThrThrGlnProTyThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 103 CATGCTGCC-----TTTTCATGTAAGAACGTCAAGTCTGGCGCGCGCGCGC 150
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyThrValAlaThrAsp 83
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 151 TTCTGTCCTCCCGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATATGCACGAACAGAT 210
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 84 IleGlyGlyMetTyArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 211 ATTGGCGGCTTACCGCCTCAAC---GCCGACGACTCATGGACCCCGCTACGGATGGG 267
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 104 Val-----GlyTrpAsnAsnTrpGlyTyAsnGlyValValSerIleAlaAla 119
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 268 ATTGCTGATAATGCGGCTGGCACACTGG-----GGCATCGACGCTGTTCGCTT 318
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyThrAsnSerTrpAsp 139
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 319 GATCGCAGCAGCATCAAAAGGTGTATGCCGAGTCGGCATGTATACGAACAGCTGGAT 378
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 379 CCGAGTAATGAGCCATCATCTCGCTCAGACCGCGCGCAACGTCCTTCACCAAC 438
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlyArgLeuAlaVal 179
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 439 TTGCGCTTCACAGTCGGGGGTAACTGTCAGGACCGGAGCCGAGAGCGTCTGGCTGTC 498
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 180 AspProAsnAsnAspAsnIleLeuTyPheGlyAlaProSerGlyLysGlyLeuTrpArg 199
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 499 GATCGCGCCACCTCCCAACATCATCTTGTGTCGCTCAGCAGACGCGCTCTGGAAG 558
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTy 219
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 559 TCTACCGAGCGGGCGTGACCTTTTCCAAAGCGCTCGCTCAGCGCCCTGGGACGTC 618
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 220 IleAlaAsnProThrAspThrThrGlyTyrglnSerAspIleGlnGlyValValTrpVal 239
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 619 CTCCCAGACCCGAGTGATTCCAAACGCTACCCCGGACCAAGCAGCCCTCATGTGGTT 678
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 240 AlaPheAspLysSerSerSerSerLeuGly-GlnAlaSerTyThrIlePheValGlyVa 259
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 679 ACGTTCCGACTCACCCAGCATCAGCAGCGGTGGGAGCCACGCTCGTATTTTGTGTCAC 738
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Qy 259 lAlaAsp 261
      |||
Db 739 GACTGAC 745
      |||

```

RESULT 4

CB905388

LOCUS

DEFINITION

jecorina cDNA clone tric074xd02, mRNA sequence.

CB905388

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

CB905388 814 bp mRNA linear EST 02-JUL-2003
tric074xd02 T.reesei mycelial culture, Version 3 april Hypocrea
jecorina cDNA clone tric074xd02, mRNA sequence.

CB905388

CB905388.1 GI:30120046

EST.

Hypocrea jecorina (anamorph: Trichoderma reesei)

Hypocrea jecorina

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; Hypocreaaceae; Hypocrea.

1 (bases 1 to 814)

Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,

Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,

Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,

Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.

Transcriptional regulation of biomass-degrading enzymes in the

Db	93	TACGGCACCGGGAATGACAATCTTTGGCGGCACACGATCTCACCAACTGGGACACGCGCCAC	152
Qy	463	GlnIleHisIleAlaProMetVallysglyLeuGluGluThrAlaValAsnAspLeuIle	482
Db	153	AATGTGTCATCAATCACTGGCAGACGGCATCGAGGAATCTCGCTCAGGACCTGGCC	212
Qy	483	SerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis	502
Db	213	TTGTCACCGCGGGAAGCGAGTATTGGCCGAGTCGAGAGACACACGGCTTCACCTTT	272
Qy	503	Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr	519
Db	273	GCCAGCAGAAACGACCTCGGGACATCGCCGACAGCGTCTGGGCAACGCCACATGGGCC	332
Qy	520	ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer	539
Db	333	ACCTCGACGAGCGTCGACTACGCGCGGAACCTCGTCAAGAGCGTGTCCGCTCGGCAAC	392
Qy	540	PheAspProSerSerGlnProAsnArgHisValAlaPheSerThrAspGlyGlyLys	559
Db	393	ACGCGCGGACGCAACAG-----GTGGCCATCTCTGTCGACGCGCGCGG	437
Qy	560	AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer	579
Db	438	ACGTGGAGCATCGACTACGCGGCCGACAGTCCATCAACGCGCGGACGCTGCTTTCG	497
Qy	580	AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla	599
Db	498	GCGACGCGGACACGATCTCTGTGTCGACCGCTCGTCGGC-----GTGACGCGC	548
Qy	600	ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg	619
Db	549	TGCGAGTTTTCAGCGGCGCTTTGCTCGCTCGAGCCTGCGCGGCGCGCTCATCGCC	608
Qy	620	SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer	639
Db	609	TGCGACAGAGACCAACAGCGTCTTCTACGCGGCTCCGGATCGACCTTTTACGTCAGC	668
Qy	640	ThrAspGlyGlyValThrPhe 646	
Db	669	AAGGACACCGGACGACGCTTC 689	
RESULT 6			
CB907625			
LOCUS			
DEFINITION		782 bp mRNA linear EST 02-JUL-2003	
		trich082xn19 T.reesei mycelial culture, Version 3 april Hypocrea	
		jecorina cDNA clone tric082xn19, mRNA sequence.	
ACCESSION		CB907625	
VERSION		CB907625.1 GI:30122283	
KEYWORDS		EST.	
SOURCE		Hypocrea jecorina (anamorph: Trichoderma reesei)	
ORGANISM		Hypocrea jecorina	
		Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	
REFERENCE		1 (bases 1 to 782)	
AUTHORS		Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,	
		Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,	
		Kelley,A.S., Weerman,H.J., Mitchell,T., Mitchinson,C.,	
		Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.	
TITLE		Transcriptional regulation of biomass-degrading enzymes in the	
JOURNAL		filamentous fungus Trichoderma reesei	
MEDLINE		J. Biol. Chem. 278 (34), 31988-31997 (2003)	
PUBMED		12788920	
COMMENT		Contact: Pamela K. Foreman	
		Genencor Intl.	
		925 Page Mill Road, Palo Alto, CA 94304, USA	
		Tel: (650) 846-7635	
		Fax: (650) 621-7817	
		Email: pforeman@genencor.com	
		Seq primer: LT-F1 primer:	
FEATURES		Location/Qualifiers	
source		1..782	

US-09-917-376-1 (1-957) x CB907625 (1-782)

Qy

423

LysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu

442

Db

108

CGCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCACCGACGACCAACCTGGCTC

167

Qy

443

TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGly

462

Db

168

TACGGCACCGGAATGACAATCTTTGGCGGCACGATCTCACCACTGGGACACGCGCCAC

227

Qy

463

GlnIleHisIleAlaProMetVallysglyLeuGluGluThrAlaValAsnAspLeuIle

482

Db

228

AATGTGTCATCAATCACTACGTCGCGACGCGATCGAGGAATCTCCGTCAGGACCTGGCC

287

Qy

483

SerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis

502

Db

288

TCTGCACCGCGGGAAGCGAGCTATTGGCGCGAGTCGAGACGACCAACGCGCTTTCACCTTT

347

Qy

503

Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr

519

Db

348

GCCAGCAGAAACGACCTCGGGACATCGCGCGACGCTCTGGGCAACGCCACATGGGCC

407

Qy

520

ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer

539

Db

408

ACCTCGACGAGCGTCGACTACGCGGGAACCTCGTCAAGAGCGTGTCCGCTCGGCAAC

467

Qy

540

PheAspProSerSerGlnProAsnArgHisValAlaPheSerThrAspGlyGlyLys

559

Db

468

ACGCGCGGACGCAACAG-----GTGGCCATCTCTGTCGACGCGCGCGC

512

Qy

560

AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer

579

Db

513

ACGTGGAGCATCGACTACGCGGCGGACACGTCCTCAACGCGCGGACGCTGTCGCTATTCG

572

Qy

580

AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla

599

Db

573

GCCAGCGGCGACACGATCTCTGTGTCGACCGCTCTGTCGGC-----GTGACGCGC

623

Qy

600

ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg

619

Db

624

TCGCAGTTCAGGGCAGCTTTGGCTCGCTCGAGCCTGCGCGGCGCGCTCATCGCC

683

Qy

620

SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer

639

Db

684

TCGGACAGAGAACCAACAGCGTCTTCTACGCGCGCTCCGGATCGACCTTTTACGTCAGC

743

Qy

640

ThrAspGlyGlyValThrPhe 646

Db

744

AAGGACACCGGACGACGCTTC 764

RESULT 7

CF882065

LOCUS

CF882065

693 bp

mRNA

linear

EST 31-OCT-2003


```

DEFINITION   trico29xol3.b11 T.reesei mycelial culture, Version 6 October 2003
ACCESSION    CF882065
VERSION      CF882065.1 GI:38136747
KEYWORDS     EST.
SOURCE       Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM     Hypocrea jecorina
              Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE    1 (bases 1 to 693)
AUTHORS      Diener, S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,
              Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,
              Ward, M. and Dean, R.A.
TITLE        Characterization of the protein processing and secretion pathways
              in a comprehensive set of expressed sequence tags from Trichoderma
              reesei
JOURNAL       FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
COMMENT      Contact: Ralph A. Dean
              Fungal Genomics Laboratory
              North Carolina State University
              Campus Box 7251, Raleigh, NC 27695, USA
              Tel: 919-513-0020
              Fax: 919-513-0024
              Email: ralph.dean@ncsu.edu
              Seq primer: IT-F1 primer.
FEATURES     Location/Qualifiers
              1..693
               /organism="Hypocrea jecorina"
               /mol_type="mRNA"
               /strain="QM6a"
               /db_xref="taxon:51453"
               /clone="trico29xol3"
               /dev_stages="mycelia"
               /clone_lib="T.reesei mycelial culture, Version 6 October
               2003"
               /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
               culture grown from 24 hrs to 6 days with varying Carbon
               and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.:      1.25e-14      Length:      693
Score:          401.50      Matches:    88
Percent Similarity: 62.30%      Conservative: 26
Best Local Similarity: 48.09%      Mismatches: 49
Query Match:    7.82%          Indels:     20
DB:             7            Gaps:       6

US-09-917-376-1 (1-957) x CF882065 (1-693)

QY 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
Db 47 TCTCGAGTCCTGGCTTGTCTGGGGCCGTC-----ATCCCTGCC 88
QY 45 HisAlaAlaThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
Db 89 CATGCTGCC-----TTTTCATGGAGAACGTCAGCTCGGGCGCGCGCGCGC 136
QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 137 TTGTCCTCCCGCATCATCTTCCATCCCAAGACAAAGCGGTAGCATATGCACGAACAGAT 196
QY 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 197 ATTGGCGGGTGTACCGCTCAAC---GCCGACGACTCATGCGCGCGCTCAGCATGGG 253
QY 104 Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 254 ATTGCTGATAATAGCGCGTGGCACAACATGG-----GGCATCGACCTTGTTCGGTT 304
QY 120 AspProIleAsnThrAsnIysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
Db 305 GATCCGCATGACGATCAAAAGGTGTATGTCGCGCACCGCGCATGTATACGAACACCTGAGAT 364

```

```

QY 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 365 CCTACTAATGAGCCCATCATTCCTCGTCAGACACCGCGCAACGTGGTCTCTCCACCAAC 424
QY 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
Db 425 TTGCGCTTTCACAGTCAGGGGTAAACATGCCAGACTCGGAGCGGAGAGCGTCTGGGTGTC 484
QY 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeu-TrpAr 199
Db 485 AATCCCTCCAACTACACATCATCTACTGTGGCGTCTGCTCAGAAACAGCCTCTCTGGAA 544
QY 199 gSerThr 201
Db 545 GTCTACG 551

RESULT 8
CE342004      748 bp      DNA      linear      GSS 26-SEP-2003
LOCUS        tigr-gss-dog-17000334095340 Dog Library Canis familiaris genomic,
DEFINITION   genomic survey sequence.
ACCESSION    CE342004
VERSION      CE342004.1 GI:36167870
KEYWORDS     GSS.
SOURCE       Canis familiaris (dog)
ORGANISM     Canis familiaris
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE    1 (bases 1 to 748)
AUTHORS      Kirsch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
              Venter, J.C.
TITLE        The dog genome: survey sequencing and comparative analysis
JOURNAL      Science 301 (5641), 1898-1903 (2003)
MEDLINE      22875432
PUBMED       14512627
COMMENT      Contact: Kirkness EF
              The Institute for Genomic Research
              Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
              Rockville, MD 20850, USA
              Tel: 301-838-0200
              Fax: 301-838-0208
              Email: ekirknes@tigr.org
              Class: shotgun.
              Location/Qualifiers
               1..748
                /organism="Canis familiaris"
                /mol_type="genomic DNA"
                /strain="Standard Poodle"
                /db_xref="taxon:9615"
                /clone_lib="Dog Library"
                /note="Site 1: BstXI; Libraries were prepared from
                peripheral blood"
ORIGIN
Alignment Scores:
Pred. No.:      3.37e-12      Length:      748
Score:          362.00      Matches:    72
Percent Similarity: 83.91%      Conservative: 1
Best Local Similarity: 82.76%      Mismatches: 14
Query Match:    7.05%          Indels:     0
DB:             9            Gaps:       0

US-09-917-376-1 (1-957) x CE342004 (1-748)

QY 782 GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer 801
Db 194 GGCCACCTTCACCTTCACCTTCCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCT 253
QY 802 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 821
Db 254 CTTTACCTTCTCTCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCA 313

```



```

FEATURES
  source      Seq primer: LT-F1 primer.
              Location/Qualifiers
                1..803
                /organism="Hypocrea jecorina"
                /mol_type="mRNA"
                /strain="QM6a"
                /db_xref="taxon:51453"
                /clone="trio16xm14"
                /dev_stage="mycelia"
                /clone_lib="T.reesei mycelial culture, Version 3 april"
                /note="vector: PREP3Y; Site 1: Not I/Sal I; Mycelial
                culture grown from 24 hrs to 6 days with varying Carbon
                and Nitrogen sources and concentrations."
ORIGIN
Alignment Scores:
Pred. No.:      3,96e-12      Length:      803
Score:          361.50      Matches:      88
Percent Similarity: 51.61%      Conservative: 40
Best Local Similarity: 35.48%      Mismatches: 103
Query Match:    7.04%      Indels: 17
DB:             6      Gaps: 6

US-09-917-376-1 (1-957) x CB898982 (1-803)
Qy      453 AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValIysGly 472
Db      67 CACAGTTTCACCACTCGGACGACGGCCACATGTGTCATTCATCAATCATCTGGCAGACGCG 126

Qy      473 LeuGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSer 492
Db      127 ATCGAGGAATTCCTCGTCAGACCTGGCTCTGCACCGCGGAAGCGAGCTATTGGCC 186

Qy      493 AlaLeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValPro 509
Db      187 GCAGTCGGAGACGACACAGCGCTTCCATTTGCCAGCAGAAACGACTCGGGACATCGCGC 246

Qy      510 SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspThrAlaGluLeu 529
Db      247 CAGACGGTCTGGGCAACCCACATCGGGCCACTTCGACGAGCGTGCATCGCGGGAAAC 306

Qy      530 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 549
Db      307 TCGGTCAAGAGGTGCTCGCGTCGGCAACACCGCGCGCACGACAG-----354

Qy      550 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 569
Db      355 ---GTGGCCATCTCTCCGACGCGCGCGACGTCGATCGATCGATCGCGCGCGACACG 411

Qy      570 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589
Db      412 TCCATGAACGGCGGCGCGGCGTGTGCGCTATTTCGGCGGACGCGCGACACGATCCTCTGTCGAC 471

Qy      590 GlyAspProGlyGlnProValValThrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 609
Db      472 GCCTGTCGCGC-----GTGCAGCGCTCGAGTTCACGGCGAGCTTTCCTCCGTC 522

Qy      610 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629
Db      523 TCGAGCTCGCGCGCGCGCGCTCATCGCTCGGACAGAACACACAGCGCTCTCTTAC 582

Qy      630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649
Db      583 GCGGCTCGGATCGACCTTTTACGTGACAGAGGACACCGCGGACGACTTC-----633

Qy      650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 667
Db      634 ACGCGCGG---CCCAAGCTGGCGGACGCGGAGGACGATCGCGGATATCGCTCTCACCG 690

Qy      668 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 687
Db      691 ACCACCGCGGACGCTGTATGTCTCGACGCGCTCGGCATATTCGCTCCACAGACTCG 750

688 GlySerSerTrpSerAlaIleThr 695
751 GGCAGCACTTTGGCCAAAGTCTCC 774

CNS028QT      791 bp      DNA      linear      GSS 01-SEP-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
245G14 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL186302
GI:7824406
GSS: genome survey sequence.
Tetraodon nigroviridis
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20396633
10835645
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
10899143
3 (bases 1 to 791)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
  source      Location/Qualifiers
                1..791
                /organism="Tetraodon nigroviridis"
                /mol_type="genomic DNA"
                /db_xref="taxon:99883"
                /clone="245G14"
                /clone_lib="G"
                /note="Genoscope sequence ID : COAG245BD07LP1-end : T7"
ORIGIN
Alignment Scores:
Pred. No.:      4,78e-12      Length:      791
Score:          360.00      Matches:      71
Percent Similarity: 84.88%      Conservative: 2
Best Local Similarity: 82.56%      Mismatches: 13
Query Match:    7.01%      Indels: 0
DB:             9      Gaps: 0

US-09-917-376-1 (1-957) x CNS028QT (1-791)
Qy      783 AlaProSerGlySerProSerProSerProSerProSerProSerProSerProSerPro 802
Db      347 AGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCT 406

Qy      803 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerSer 822
```


Db 626 CCGCCAGCGGATGGAGTCATTTGCAAACTGTAGCTGCTTCTTTCATTTGTCCCACTGAGG 685
Qy 894 GlnVal -----ValAsnThrGlySerSerValAspLeu 905
Db 686 CAAATTTCGATGCAGATTGCACCGGTATGAATGCACACACACACTTGATTTT 745
Qy 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925
Db 746 GGGTCATTGTGTATTTTATTTT-----TCTACAGTTTATAT 784
Qy 926 AsnCysAsp 928
Db 785 TG-TGTGAT 792

Search completed: October 5, 2005, 21:09:16
Job time : 9216.32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 5, 2005, 07:47:30 ; Search time 455.52 Seconds
(without alignment)
3437.649 Million cell updates/sec

Title: US-09-917-376-1
Perfect score: 5135
Sequence: 1 MDRSENRLTWRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q-/cgn2_1/USPTO spoel/US09917376/runat 04102005 164328 26925/app query.fasta_1.2524
-DB=issued Patents_NA -QFWT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0 1 -ILOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CGN 1.1 516 @runat 04102005 164328 26925 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTC COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	409	8.0	11707	3	US-09-136-574A-1
2	358	7.0	19383	4	US-09-949-016-16031
3	345	6.7	18596	3	US-09-318-448-11
4	345	6.7	18597	4	US-09-962-665-8
5	345	6.7	18597	4	US-09-963-333-8
6	345	6.7	18597	4	US-09-962-677-8
7	334	6.5	6416	3	US-09-136-574A-2
8	319.5	6.2	8211	4	US-09-252-991A-13656
9	296.5	5.8	4767	4	US-09-410-551B-28
10	296.5	5.8	4767	4	US-09-940-316B-28
11	293	5.7	4818	4	US-09-410-551B-32
12	293	5.7	4818	4	US-09-940-316B-32

C	13	288.5	5.6	4131	4	US-09-252-991A-13773	Sequence 13773, A
	14	288.5	5.6	4571	4	US-09-410-551B-18	Sequence 18, Appl
	15	288.5	5.6	4571	4	US-09-940-316B-18	Sequence 18, Appl
	16	288	5.6	288	4	US-09-119-507B-111	Sequence 111, App
	17	288	5.6	288	4	US-09-547-693-111	Sequence 111, App
	18	282.5	5.5	441529	3	US-09-103-840A-1	Sequence 1, Appli
	19	281.5	5.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	20	280	5.5	3129	4	US-09-252-991A-13873	Sequence 13873, A
	21	279	5.4	4547	4	US-09-410-551B-22	Sequence 22, Appl
	22	279	5.4	4547	4	US-09-940-316B-22	Sequence 22, Appl
	23	273	5.3	4674	4	US-09-410-551B-26	Sequence 26, Appl
	24	273	5.3	4674	4	US-09-940-316B-26	Sequence 26, Appl
	25	269	5.2	4478	4	US-09-410-551B-16	Sequence 16, Appl
	26	269	5.2	4478	4	US-09-940-316B-16	Sequence 16, Appl
	27	268.5	5.2	31467	4	US-09-949-016-13134	Sequence 13134, A
	28	268.5	5.2	31868	4	US-09-949-016-11907	Sequence 11907, A
	29	268	5.2	31391	4	US-09-949-016-14319	Sequence 14319, A
	30	266	5.2	4188	4	US-09-252-991A-13774	Sequence 13774, A
	31	266	5.2	77536	4	US-09-410-551B-1	Sequence 1, Appli
	32	266	5.2	77536	4	US-09-940-316B-1	Sequence 1, Appli
	33	265.5	5.2	77536	4	US-09-410-551B-1	Sequence 1, Appli
	34	265.5	5.2	77536	4	US-09-940-316B-1	Sequence 1, Appli
	35	265	5.2	4737	4	US-09-410-551B-30	Sequence 30, Appl
	36	265	5.2	4737	4	US-09-940-316B-30	Sequence 30, Appl
	37	262	5.1	150394	4	US-09-949-016-13042	Sequence 13042, A
	38	260	5.1	3337	2	US-08-072-610-1	Sequence 1, Appli
	39	260	5.1	3337	2	US-08-719-822B-1	Sequence 1, Appli
	40	260	5.1	3337	3	US-09-092-458-1	Sequence 1, Appli
	41	260	5.1	3337	4	US-08-719-821C-1	Sequence 1, Appli
	42	260	5.1	46253	4	US-09-949-016-11890	Sequence 11890, A
	43	260	5.1	46257	4	US-09-949-016-13711	Sequence 13711, A
	44	258	5.0	4725	4	US-09-410-551B-24	Sequence 24, Appl
	45	258	5.0	4725	4	US-09-940-316B-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-09-136-574A-1
; Sequence 1, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
 REFERENCE/DOCKET NUMBER: 1997US001/CIP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-540-9200
 TELEFAX: 215-540-5818
 TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 11707 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-136-574A-1

Alignment Scores:
 Pred. No.: 1,52e-11 Length: 11707
 Score: 409.00 Matches: 263
 Percent Similarity: 32.37% Conservative: 163
 Best Local Similarity: 19.98% Mismatches: 386
 Query Match: 7.96% Indels: 504
 Gaps: 58

US-09-917-376-1 (1-957) x US-09-136-574A-1 (1-11707)

```

Qy 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33
Db 735 AAGAGGGTAATTCATTCCTT---TCCTTATTTGTTTTTAAATAAACACCGTTGAGGT 791

Qy 34 ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThr 53
Db 792 ACTTTG-----ATATTTTCATCAGCAAGCAAAAGCAGCAGCA-----TATACT 833

Qy 54 TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGly 73
Db 834 -----GTTGATTTTGAAGGT 848

Qy 74 AlaProGlyIleLeuTyr-----ValArgThrAspIleGlyGly 86
Db 849 GCTGATATTTATCTTACTTTGCTTANGGAAATCGACATAGCAGTGTGACATGGGCAAT 908

Qy 87 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 106
Db 908 ----- 908

Qy 107 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126
Db 909 -----GCATATAATGGTAAAGTAGTGTGTCAGGGGTGCA-----AATAGAAGTTCA 953

Qy 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146
Db 954 ATATGGGATGGAGTTCAGTT-----GACGTTAAA 983

Qy 147 ArgSerAspGlnGlyAlaThrTrpGlnIleThr----- 158
Db 984 AACATATGAACAATGGAAACACATGGGTAGTTTCAGCGTATGTATTAACATAGCTACCAG 1043

Qy 159 ---ProLeuProPheLysLeu-----GlyGlyAsnMetProGlyArg 171
Db 1044 AAGCGGTTGCTATTTGGTATCTCAGCGGTTTACGACGATGGGAAGTGGGTTAAGAGTACT 1103

Qy 172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191
Db 1104 CTCATAGGTAGGTTGTGGCTATT---CCAAAT----- 1133

Qy 192 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211
Db 1134 -----TATTGGAAGAAAATT--- 1148

Qy 212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 231
Db 1149 -----GTTGGTAAATGGACTCCAAATATATAGCAATGTGACGAAATTTGTTAATT 1196

```

```

Qy 232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 251
Db 1197 GTAATACACACAATTTAGAAAGCGAAGTAGATTATAATGTTGAC----- 1241

Qy 252 SerLysThrIlePheValGlyValAlaAspProAsn-----AsnProValPhe 267
Db 1242 -----TATATCCAAATAATGGATGATTAATAGTTTACCTATCAATGAGTACACA 1289

Qy 268 TrpSerArg-----AspGlyGlyAlaThr-----TrpGlnAla-----Val 279
Db 1290 TTTTCAAGTGGATTTGAAAGTGGCACTACCGAGGGTTGGCAGGCAAGGGGAGCGGTGTT 1349

Qy 280 ProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisVal 299
Db 1350 ACAGTAAACACAGATAGCGTTGTCATATATAGTGGCAAGTATAGT----- 1394

Qy 300 LeuTyrIleAla-----ThrSerAsnThrGlyGly-----ProTyrAspGly--- 313
Db 1395 TTGTACGTCACTGAGGAAGACGTCAAATTTGGCATGGTGCACAGATTCGGGTAGATACAATT 1454

Qy 314 ---SerSerGlyAspValTrpLysPheSerVal-----ThrSerGlyThrTrp 328
Db 1455 TTGGAACAGGGTAAAGTGTATAAATAAGTGTGTTGGGTTTATCAGAACACGTGGTTCAACT 1514

Qy 329 ThrArgIleSerProValProSerThrAspThrAlaAsnAsp---TyrPheGlyTyrSer 347
Db 1515 CAAAATAATGTCATTAACCTATGCAAGAACATTTGCTACAGATCTCTCAACACAGCTAGAA 1574

Qy 348 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 367
Db 1575 AATCTGATATATAACAGGGATGTACCGAGTAATACCTGGGTTGAGCTGAGTGAAGAGTAC 1634

Qy 368 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387
Db 1635 -----TCAATTCCTGCTGGTGTACAGTTACAGTACGAGTTGTTG 1670

Qy 388 AspTrpThrSerTyrProAsnArgSerLeuArgTyr-----ValLeu 401
Db 1671 CTTTATGTTGAGCGCAAAATGCAAAATTTGGCTTCTGGTGTGATGATTAAGATTAT 1730

Qy 402 AspIleSer-----AlaGluPro----- 407
Db 1731 GATTATCCAAGTGTGCTGAACCTGAATGGAGATACCATCTTTGTATAGAAAAGTATAGA 1790

Qy 407 ----- 407

Db 1791 GATTATTTCAAAGTAGGAGTAGCTTTGTCTTTACAAAAGCATTCGCTCTGTATACAGAAAAG 1850

Qy 407 ----- 407

Db 1851 AAGATGGTTTGAAGCATTTCAATAGTATTACTGTCAGGGAACGAATGAACCATCAGAG 1910

Qy 408 -----TrpLeuThrPheGly 412
Db 1911 TTACTTGTGATGAAAATACTTACAACTTTAGCAAGCAGACGAAATTTGTAAATTTTGCA 1970

Qy 413 ValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAla----- 430
Db 1971 ACAAGTAAACAACATTTGCCATCAGAGGTATACACTGTTTGGCATGAGCAACACCCGAC 2030

Qy 430 ----- 430

Db 2031 TGGTTTTTCAAGGACACAAATGGAATATACGTTGAGCAAGGATGCAATGCTTAAGCAGATTA 2090

Qy 430 ----- 430

Db 2091 AACAGTATATTTATACGGTAGTGGGAAGATATAAAGGGAAGGTTTATGATCGATGGATGTG 2150

Qy 431 -----MetAlaIleAspProPheAsnSerAspArgMetLeu----- 442
Db 2151 GTAAATGAAGCAATAGATCAAAAGTCAAGGTGTGGATTCAGGAGATCTAACTGGTACAAC 2210

Qy 443 -----TyrGlyThrGlyAlaThrLeuTyrAla----- 451

```


LOCATION: 15042, 15546, 15770
OTHER INFORMATION: n = c or t
NAME/KEY: misc_feature
LOCATION: 1322, 1688
OTHER INFORMATION: n = c or g
NAME/KEY: misc_feature
LOCATION: 2594, 11293, 16199, 16203
OTHER INFORMATION: n = g or t
NAME/KEY: misc_feature
LOCATION: 3619
OTHER INFORMATION: n = a or t
NAME/KEY: misc_feature
LOCATION: 14547
OTHER INFORMATION: nucleotide in position 14547 is t, or absent

US-09-962-665-8

Alignment Scores:
Pred. No.: 4.92e-08 Length: 18597
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13
Query Match: 6.72% Indels: 1
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-962-665-8 (1-18597)

QY 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db 10388 TCACCATCACCATTCTCCATCCATCACCATCACCATCACCATCACCATCACCACCA 10329
QY 803 SerProSerProSerProSerProSerProSerProSerProSerProSerSer 822
Db 10328 TCACCATCTCCATCACCATCACCATCACCATCTCCATCACCATCACCATCACCACCA 10269
QY 823 SerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
Db 10268 TCACCATCACCATCACCATCTCCATCACCATCACCATCACCATCTCCATCACCACCA 10209
QY 843 SerProSerProSerProSerProSerProSerProSerProSerProSerSer 862
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10149
QY 862 othrProSerSerPro 868
Db 10148 ATCTCCATCACCATCACCACCA 10130

RESULT 5

US-09-963-333-8/c
Sequence 8, Application US/09963333
Patent No. 6664062
GENERAL INFORMATION:
APPLICANT: Stanton, Jr., Vincent P.
TITLE OF INVENTION: THYMIDINE SYNTHASE GENE SEQUENCE VARIANCES
TITLE OF INVENTION: HAVING UTILITY IN DETERMINING THE TREATMENT
TITLE OF INVENTION: OF DISEASE
FILE REFERENCE: 11926-015002
CURRENT APPLICATION NUMBER: US/09/963,333
CURRENT FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: 09/658,659
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 09/596,033
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 09/357,743
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 09/357,024
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: 60/093,484
PRIOR FILING DATE: 1998-07-20
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 18597
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 701, 13751
OTHER INFORMATION: n = c or a
NAME/KEY: misc_feature
LOCATION: 716, 1293, 2401, 2429, 2618, 3083, 3125, 3635, 4256, 4898,
LOCATION: 5062, 5167, 11069, 13298, 14479, 14730, 14796, 15344, 15450,
LOCATION: 15503, 15590, 15840, 16149
OTHER INFORMATION: n = a or g
NAME/KEY: misc_feature
LOCATION: 732, 1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,
LOCATION: 12598, 13171, 13645, 13782, 13806, 13813, 14586, 14788,
LOCATION: 15042, 15546, 15770
OTHER INFORMATION: n = c or t
NAME/KEY: misc_feature
LOCATION: 1322, 1688
OTHER INFORMATION: n = c or g
NAME/KEY: misc_feature
LOCATION: 2594, 11293, 16199, 16203
OTHER INFORMATION: n = g or t
NAME/KEY: misc_feature
LOCATION: 3619
OTHER INFORMATION: n = a or t
NAME/KEY: misc_feature
LOCATION: 14547
OTHER INFORMATION: nucleotide in position 14547 is t, or absent

US-09-963-333-8

Alignment Scores:
Pred. No.: 4.92e-08 Length: 18597
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13
Query Match: 6.72% Indels: 1
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-963-333-8 (1-18597)

QY 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db 10388 TCACCATCACCATTCTCCATCCATCACCATCACCATCACCATCACCATCACCACCA 10329
QY 803 SerProSerProSerProSerProSerProSerProSerProSerProSerSer 822
Db 10328 TCACCATCTCCATCACCATCACCATCACCATCTCCATCACCATCACCATCTCCATCACCACCA 10269
QY 823 SerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
Db 10268 TCACCATCACCATCACCATCTCCATCACCATCACCATCACCATCTCCATCACCACCA 10209
QY 843 SerProSerProSerProSerProSerProSerProSerProSerProSerSer 862
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10149
QY 862 othrProSerSerPro 868
Db 10148 ATCTCCATCACCATCACCACCA 10130

RESULT 6

US-09-962-677-8/c
Sequence 8, Application US/09962677
Patent No. 6759200
GENERAL INFORMATION:
APPLICANT: Stanton, Jr., Vincent P.
TITLE OF INVENTION: THYMIDINE PHOSPHORYLASE GENE SEQUENCE
TITLE OF INVENTION: VARIANCES HAVING UTILITY IN DETERMINING
TITLE OF INVENTION: THE TREATMENT OF DISEASE
FILE REFERENCE: 11926-015003
CURRENT APPLICATION NUMBER: US/09/962,677
CURRENT FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: 09/658,659
PRIOR FILING DATE: 2000-09-08


```

; PRIOR APPLICATION NUMBER: 09/596,033
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 09/357,743
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 09/357,024
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: 60/093,484
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 18597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 701, 13751
; OTHER INFORMATION: n = c or a
; NAME/KEY: misc_feature
; LOCATION: 716, 1293, 2401, 2429, 2618, 3083, 3125, 3635, 4256, 4898,
; LOCATION: 5062, 5167, 11069, 13298, 14479, 14730, 14796, 15344, 15450,
; LOCATION: 15503, 15590, 15840, 16149
; OTHER INFORMATION: n = a or g
; NAME/KEY: misc_feature
; LOCATION: 732, 1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,
; LOCATION: 12598, 13171, 13645, 13782, 13806, 13813, 14586, 14788,
; LOCATION: 15042, 15546, 15770
; OTHER INFORMATION: n = c or t
; NAME/KEY: misc_feature
; LOCATION: 1322, 1688
; OTHER INFORMATION: n = c or g
; NAME/KEY: misc_feature
; LOCATION: 2594, 11293, 16199, 16203
; OTHER INFORMATION: n = g or t
; NAME/KEY: misc_feature
; LOCATION: 3619
; OTHER INFORMATION: n = a or t
; NAME/KEY: misc_feature
; LOCATION: 14547
; OTHER INFORMATION: nucleotide in position 14547 is t, or absent
US-09-962-677-8

Alignment Scores:
Pred. No.: 4,92e-08 Length: 18597
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13
Query Match: 6.72% Indels: 1
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-962-677-8 (1-18597)
QY 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db 10388 TCACCATCACCATTCCATCTCCATCACCATCACCATCACCATCACCATCACCATCACC 10329
QY 803 SerProSerProSerSerProSerProSerProSerProSerProSerProSerSer 822
Db 10328 TCACCATCTCCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACC 10269
QY 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
Db 10268 TCACCATCACCATCACCATCTCCATCACCATCACCATCACCATCACCATCACCATCACC 10209
QY 843 SerProSerProSer-SerSerProSerProSerSerProSerProSerProSerPro 862
Db 10208 TCACATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACC 10149
QY 862 cThrProSerSerPro 868
Db 10148 ATCTCCATCACCATCACCACCA 10130
RESULT 7

US-09-136-574A-2
; Sequence 2, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; ANDERSON, Paige
; GIBBS, Moreland
; BERGQUIST, Peter
; DANIELS, Roy
; MORGAN, Hugh W.
; WILLIAMS, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESS: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2

Alignment Scores:
Pred. No.: 5,27e-08 Length: 6416
Score: 334.00 Matches: 219
Percent Similarity: 33.27% Conservative: 125
Best Local Similarity: 21.18% Mismatches: 314
Query Match: 6.50% Indels: 376
DB: 3 Gaps: 56

US-09-917-376-1 (1-957) x US-09-136-574A-2 (1-6416)
QY 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIle 37
Db 1189 TCTGTAGTGGCAGAAACA-----GCTGCATCACTTCTGGTGGCTTCAGTT 1233
QY 38 AlalIleThrAlaSerProAlaHisAlaAla----- 47
Db 1234 GTAATTAAGGAAGAATAATCTCAGAAAGCAGCTTCTATCCACATGCCAAAGACCTG 1293
QY 48 -----ThrThrGlnPro-----TyrThrTrpSerAsnValAlaIle 59
Db 1294 TTTGAATTGGCGATACCAACAAGAGTGTATCGGGGTACTACTGCTGCACAGGTTTCTAC 1353
QY 60 GlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyr 79
```


Db 1354 ACATCGGTGTTTATTGATGACCTTGGATGG-----GCTGCTGATGGCTTAT 1404
QY 80 ValArgThrAsp-----
Db 1405 ATTGCGACAAATGACAGTAGTTATTGCGAAAGCTGAAGAGTTGATGTCAGAAATATGCT 1464
QY 85 GlyGlyMetTyrArgTTP-----AspAlaAlaAsnGlyArgTTPilePro 99
Db 1465 AATGGTACTAATACATGACACAAATGCTGGGATGATGTTCCGATGGAACATTGATCATG 1524
QY 100 LeuLeuAspTTPValGlyTTPAsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAla 119
Db 1525 CTTCGAAGATTACAGGAAGAGTTA-----TATAAGGAGCTGTGGAAGAACTTA 1578
QY 120 AspProIleAsnThrAsnLysValTTPAlaAlaValGlyMet-----TyrThrAsnSerTTP 138
Db 1579 GACCATTTGGACTGACAGAAATTACGTATACGCCGAAGGGATGGCATATCTGACAGGATGG 1638
QY 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGly----- 152
Db 1639 -----GGTTTCATTAGATATCGCACACACAGCTGCATTTTTTAGCATGTGCTAT 1686
QY 153 AlaThrTTPGln-----IleThr 158
Db 1687 GCAGACTGGTCAGGTGCGATTCGACAAAGAACCAATATTTGCACTTTCGCAAAAGC 1746
QY 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArg-----GlyMetGly 174
Db 1747 CAGATTGACTATGCACTGGTTCACCA-----GGTAGAAGTTTGTAGTAGGATTGGC 1800
QY 175 GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGly 194
Db 1801 ACCAATTATCCAAACATCCGCATCACAGGAAT----- 1833
QY 195 LysGlyLeuTTPArgSerThrAspSerGlyAlaThrTTPSerGlnMetThrAsnPhePro 214
Db 1834 -----CGCATAGTTTCATGGCTAAACAGCATGAAATACCA 1869
QY 215 Asp-----ValGlyThrTyrIleAlaAsnProThrAspThrThrGly 228
Db 1870 GAGTATCAGACACATATATATGAGCACTGGTGTGGTCTCTGGTAGTGATGATAGT 1929
QY 229 TyrGlnSerAspIleGlnGlyValValTTP-----ValAlaPheAspLysSerSer 246
Db 1930 TATAATGATGACATTACCGATTATGTACAAATGAGGTTGCTCGCATTAATAATGCTGGA 1989
QY 247 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 266
Db 1990 ATTGTTGGTGCATCGCAAG----- 2010
QY 267 PheTTPSerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe 286
Db 2011 -----ATGTACAGTTATATGAGTGGAACCTATTGAT--- 2043
QY 287 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr----- 304
Db 2044 ---GATTTTAAACCAATTGAAACACCACCAAAATGATGAAATTTTCTGAAATCAAAATTT 2100
QY 305 SerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTTPLysPheSerValThr 324
Db 2101 GCGAATTTCAGGGTCCA-----AATTATACCGAAGTAATTTCTATATCTATAAT 2151
QY 325 SerGlyThrTTPThrArgIleSerProValProSerThrAspThrAlaAsn---AspTyr 343
Db 2152 CGAACAGGATGG-----CCACCAAGGGTAACATGATAACTAGTTTAATAT 2199
QY 344 Phe-----GlyTyrSerGlyLeuThrIleAspArgGln 354
Db 2200 TTATAGACCTAACCAATTAATCCAGGCAGGTAATCG----- 2238
QY 355 HisProAsnThrIleMetValAlaThrGlnIleSerTTPTrpProAspThrIleIlePhe 374
Db 1000

2239 ---CCTGATGTTGTCAAAAGTTGCACACA-----TAC 2265
QY 375 ArgSerThrAspGlyGlyAlaThrTTPThrArgIleTTPAspTTPThrSerTyrProAsn 394
Db 2266 TACATCGAAGAGGTAAATATTAGCGTCTCTTACGTTATGGAC-----AAAAAT 2313
QY 395 ArgSerLeuArgTyrValLeu---AspIleSerAlaGluProTTPLeuThrPheGlyVal 413
Db 2314 AGGAATATATACTATGTTCTTGTGATTTTAGTGAACCAAG-----ATA 2358
QY 414 GlnProAsnProProValProSerProLysLeuGlyTTP-----MetAspGlu 429
Db 2359 TATCTCTGGCGGTGAAGTTGAACACAAAAGCAGGCTCAATTTAAATATATCTTCCGAC 2418
QY 430 AlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeu 449
Db 2419 GGGTATCCATGGGATCTTACCAAT---GATCCTCATATAAGGATTAACCATGTCATTA 2475
QY 450 TyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIleAlaProMet 469
Db 2476 GAAAGAAATAATATATATTCGCCCATATGATAATAAT---CTG 2517
QY 470 ValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPro----- 484
Db 2518 GTATGGGTTTAGAGCGCGGTGCGCAACATCCACACCTGCACCAACATCAACACCAACA 2577
QY 485 -----ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 502
Db 2578 CCAACCCCGCCCAACACACACAGTGACAGCA----- 2610
QY 503 AlaAspValThrAlaValProSerThrIlePheThr-----SerProValPheThr 519
Db 2611 -----ACGCCGAGCGGACTCTACACCGACGCGGGTCACT---GGT 2655
QY 520 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleLeValArgAlaGlySer 539
Db 2656 ACGGGAAGTGGTGAAGGTACTGTACAAGAACAAATGAGACAAATGCGGACACAGTTCT 2715
QY 540 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 559
Db 2716 ATAAGCGCG----- 2724
QY 560 AsnTTPPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 579
Db 2725 ---TGTTTAAAG-----ATAGTGAATCGAGC-----AGCAGCACT 2757
QY 580 AlaAspGlySerArgPhe-----ValTTP-----AlaProGlyAspProGlyGln 594
Db 2758 GTTGATCTTAGCAGGTTAAGATAAGATACTGTTACACAGTGGATGGTGACACAGCCACAG 2817
QY 595 ProValValTyrAlaValGlyPheGlyAsnSerTTPAlaAlaSerGlnGlyValProAla 614
Db 2818 AGTGGCGTATGT-----GACTGG----- 2835
QY 615 AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGly 634
Db 2836 ---GCACAGATAGGGGCAAGCAATGTGACA---TTCATTTTGTGAACCTTAGACGCGGA 2889
QY 635 ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 654
Db 2890 GTG-----AGTGGAGCGGATTTATAC---CTGGAGGTAGGATTT--- 2925
QY 655 SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlu---GlyAspLeu 673
Db 2926 AGCAGTGGAGCTGGCAGGTTGTCAG-----CCTGTAAGGACACAGGGGATATA 2973
QY 674 TTPLeuAlaAlaSerSerGlyLeuTyrHisThrAsnGlyGlySerSerTTPSerAla 693
Db 2974 CAGGTAGGTTTAAACAGAAATGATGAGCAATTACATCCGACAGCAGCTGTGTCATCG 3033
QY 694 IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSer 713
Db 3034 TTCAGAGCATGACGAAT-----TATGAGAGAAATGCGAAGGTGACGCTG 3078


```
QY 714 TyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAsp 733
DB 3079 TAT-----GTAGATGGT-----3090
QY 734 AspCysGlyThrThrTrpValLeuLeuAsnAspGlnHisGlnTyrGlyAsnTrpGly 753
DB 3091-----GTTCCTGGTA-----TGGGGG 3105
QY 754 GlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArg 773
DB 3106 CAG-----3108
QY 774 GlyIleValTyrGlyAspIleGlyGlyAlaProSerGlySerProSerProSerValSer 793
DB 3109-----GAGCCGGAGGAGCG-----3123
QY 794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer 813
DB 3124-----ACACTGCACCGACAGCACACAGCAACCAACG 3156
QY 814 ProSerProSerSerProSerSerProSerProSerProSerProSerProSer 833
DB 3157 CCAACTCCGACAGCAACCCCAACACCTACACCTACACCGACCCCGACACCGACAGTGAGT 3216
QY 834 ProSerArgSerProSerProSerAlaSerProSerProSerProSerProSer 853
DB 3217 GCAACGCCAACACCGGACCGGACATCACCCTAGGTGGCAGTTACTGGACCGCGAGT 3276
QY 854 SerSerProSerSerProSerProSerProSerProSerProSerProSerValSerGlyGlyVal 873
DB 3277 GAGAGT-----TACGGTCGCGTG 3294
QY 874 LysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 893
DB 3295 AAGGTATGGTATGCAATGGCAATTAGCAGCCGACCAATGTATTGAATCCTAAAGATA 3354
QY 894 GlnValValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrp 913
DB 3355 AAGATAGAGAATGTTGGGACGACACCGGTAGATCTTAGCAGGGTGAAGGTAAAGTACTGG 3414
QY 914 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaIleGly 933
DB 3415 TACAGTAGATGTAGGCGACACAGAGTGT-----3447
QY 934 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAla 947
DB 3448-----AGGTAGCGAGCAGCATAAATCCTGCG 3474
```

RESULT 8

```
US-09-252-991A-13656
; Sequence 13656, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13656
; LENGTH: 8211
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13656
```

Alignment Scores:

Pred. No.: 3.87e-07 Length: 8211

```
Score: 319.50 Matches: 290
Percent Similarity: 32.56% Conservative: 133
Best Local Similarity: 22.33% Mismatches: 449
Query Match: 6.22% Indels: 432
DB: 4 Gaps: 69

US-09-917-376-1 (1-957) x US-09-252-991A-13656 (1-8211)
QY 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIle 37
DB 3326 GCAATGGCAGCAGCCTCAGCGCACTCGGAAACCGGGAGCAGCGGTATCTCACCAGC 3385
QY 38 AlaIleThr-----AlaSerProAlaHisAlaAlaThrThrGlnPro-----51
DB 3386 GCAACGGCAATCCGATCGCGAGGTCAACCGCGAGCGGAGCGCAACTGGACCTACACCC 3445
QY 52 -----TyrThrTrpSerAsnVal-----AlaIleGly 60
DB 3446 CGTCCACGCCGATCGCCCAACGGTACTGTGGTC-AACGTGTGGCCGAGGACGCCGCGGT 3504
QY 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
DB 3505 AACAGCAGCCCGCGGCGAGGTGACCTGCTGATTCAGCGCGCGCGCGCGCGCGGTGAT 3564
QY 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
DB 3565 AACCCGACACCGGCGTGTCTATCAGCGGACCGCGCGGCGCGGTGCCACCGGACCTC 3624
QY 101 LeuAspTrpValGlyTrpAsn-----AsnTrp 109
DB 3625 ACCGAT---GCCGGCGCAACCCGATAGGCGAGGTCAACCGCGGACCGCGCAACTGG 3681
QY 110 GlyTyr-----AsnGly-----ValSerIleAlaAlaAsp 120
DB 3682 AGCTTCACCGCGGCGACCGCGCGCCCAACGGCACTGATCGTCCCGCGCACCGCAC 3741
QY 121 Profile---AsnThrAsnLysValTrpAla-----AlaValGlyMetTyrThr 135
DB 3742 CGACCGGCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3801
QY 136 AsnSerTrpAspProAsnAspGlyAlaIleLeu-----146
DB 3802 CCGGTGATCGATCCGAGCAACCGCACCATCAGCGGCGACCGCGGAGGCGCGGCGCAAG 3861
QY 147 -----ArgSerSerAspGlnGlyAla 153
DB 3862 GTGATCCTCACCGCGCAACCGCACCATCAGCGGCGGCGGCGGCGGCGGCGGCGG 3921
QY 154 ThrTrpGlnIleThrPro---LeuProPheLysLeuGly-----165
DB 3922 AACTGGACCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3981
QY 166 -----GlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 181
DB 3982 CAGGACCTCGCGGCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4041
QY 182 Asn-----AsnAspAsnIleLeuTyrPheGlyAlaProSerGly 194
DB 4042 AACACGCGCTGTGGTCAATCCGAGCAACCGCAACCTGCTCAACCGGTACCGCGGCGG 4101
QY 195 LysGlyLeu-----TrpArgSerThrAspSer 203
DB 4102 AGCACCGTGACCTTACCGGCAACCGGCAACCGGATCGGCGGCGGCGGCGGCGGCG 4161
QY 204 GlyAlaThrTrp-----SerGlnMetThrAsnPheProAspValGlyThr 218
DB 4162 AGCGGCAACTGGAGCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4209
QY 219 TyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal 237
DB 4210 GTGGTCAACGTCACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4260
QY 238 TrpValAlaPheAspLysSerSerSerSerLeuGlyGln-----AlaSerLysThr 254
```



```
QY 789 rProSerVal-----SerProSerAlaSer-----ProSerLeu----- 800
Db 6367 GCCACGGCAGTGCAACTGACCTTACCCCGCGCAGCGCGTGGCCAAACGGTACGGTG 6426
QY 801 ----SerProSerPro----SerProSerSerProSerProSerPro----- 816
Db 6427 ATCAACGGCTCGCGAAGACCGCGCGGCAACGCCAGCGGTCCGGCCACCAACGGTG 6486
QY 817 -----SerSerSerProSerSerSerProSerProSerPr 826
Db 6487 GACTCGGTGGCGCGCTCGCTCGCTGCTGAGCATCAGCGCGCAGCGCGCTGCTGACC 6546
QY 826 oSerPro-----SerProSerProSerProSerAr 836
Db 6547 GGCACCGCGCAGCGCAACAGTCAGTGCCTGATCGTTCACGCGCAGCGCAACCCG 6606
QY 836 gSerProSerProSerAlaSerProSerSer-----Se 849
Db 6607 ATCAGGTACCGCTCGACGGCGCGGCAACTTACGCTGCCGTCGCGCGCGCTGATC 6666
QY 849 rProSer-----ProSerSerProSerSerSe 859
Db 6667 ACCGCGCAACTGATCGCGGGGTTGCCGTGGACGCGCGCGCAACGTTCAGCGGGCGGCC 6726
QY 859 rProSer-----ProThrPr 864
Db 6727 ACCATCAACGCGCGGACCTGGCGCGCGCCAGCATCAGCGTCCCGAAGCGCGCATACC 6786
QY 864 oSerSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaPr 884
Db 6787 TGATCAACGCGCGGAGATCGGGACGCGC-ATCCAGTGC----- 6825
QY 884 oGlyAspAsnGlnLysProGlyLeuGlnValAlaAsnThrGlySerSerValAs 904
Db 6826 ----GATGTAGCGGTCCGTCGACCATCAGCTGCGCGAAGCGCGCATACC 6858
QY 904 pLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuVa 924
Db 6859 ----CAGGTGTGTACGGTCAAGTTC--GCGCGGCAAGACGCTACGAGCGCGAGGTGAG 6911
QY 924 lTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPhe----- 941
Db 6912 CCATACCTCACCAGCGCGGACATCGCGCGCGCAACCTGACCCCTGACCTGACGCGCTCC 6971
QY 942 -GlySerValAsnPro-----AlaThrProThrAlaAsp 952
Db 6972 CGCGCGCATGGCGCGTTCGCGAGGGCGCGCTCGACCGCTACCGCGCGAC 7020
```

RESULT 9

```
US-09-410-551B-28
; Sequence 28, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 28
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4766)
US-09-410-551B-28
```

Alignment Scores:

```
Pred. No.: 3,11e-06 Length: 4767
Score: 296.50 Matches: 233
Percent Similarity: 35.16% Conservative: 101
Best Local Similarity: 24.53% Mismatches: 397
Query Match: 5.77% Indels: 223
DB: 4 Gaps: 46
```

US-09-917-376-1 (1-957) x US-09-410-551B-28 (1-4767)

```
QY 9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28
Db 1827 CTGACCGAAACACGAAAGACCGGCTGCGCGCTACTACCTGGCGCGCTGCGCGCGGATATA 1886
QY 29 AlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThr 48
Db 1887 CGGGGTGTGGCATCGACGCTG-----GCGGTGACACACGGTTCGGT-GTTCGAGCACCGCGC 1939
QY 49 ThrGlnProTyr----- 52
Db 1940 CGTACTCTCTGGAGATGACACCGTCAACCGCACCGCGGTGACCGACCCAGGATCGTGTT 1999
QY 53 -----ThrTrpSerAsnValAlaIleGlyGlyGly-----PheVal 65
Db 2000 TGTCTTTCCCGGCGAGGGGTGCGTGGCTGGGATGGGCGAGTGCATCTGCGCATTCGTC 2059
QY 66 AspGlyIleValPheAsnGluGly-----AlaProGlyIleLeuTyrValArgThrAsp 83
Db 2060 GGTGTGTTCGCGCAGCGGATGCGCGAGTGTGCGCGCGGTTCGCGAGTTCGT---GGA 2116
QY 84 lIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 2117 CTGGGATCTGTACGGTTCGTGATGATCGGCGGTGG-TGGACCGGGTGTATG---TGG 2172
QY 104 ValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAlaProIleAsn 123
Db 2173 TCCAGCCC-GCTTCTCGGCGGATG-----ATGTTTCTCTGCGCGC----- 2213
QY 124 ThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSer---TrpAspProAsnAsp 142
Db 2214 -----GTGTGGCAGCGCGCGGTGTGCGCGCGATCGCGGTGATCGCGCATTCGCAG 2264
QY 143 GlyAlaIleLeuArgSerSerAspGlnGlyAlaAlaThrTrpGln----- 156
Db 2265 GTTGAGATCGCGCAGCTTGTGTGGCGGTGCGGTGTCTACTACCGATGCGCGCGCGATC 2324
QY 157 lIleThrProLeuProPheLysLeuGlyGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176
Db 2325 GTGACCTTGGCAGCGCAGCGATCGCGCGGGCTGCGCGCGCGCGCGCGCGATGCGATCC 2384
QY 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196
Db 2385 GTCGCCCTG---CCCGCGCAGGATGTGAGCTGTGTCGACGCGGCGC----- 2426
QY 197 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 216
Db 2427 ---TGGATCGCGCGCCACAAACCGCGCGCGCTCACCCTGTGATCGCGGCGACCCCGGAAGCG 2483
QY 217 GlyThrTyrIleAlaAsnProThrAspThrThrGly----- 228
Db 2484 GTCGACCATGTCTCTACCGCTCATGAGGCACAGGGGTGCGGTGCGCGCGATCACCGTC 2543
```


QY 229 ---TyrGlnSerAspIleGlnGlyValValTrpVal-----AlaPheAspLysSer 244
DB 2544 GACTATGCTCGCACACCCCGACGCTGAGCTGCTCCGCGAGCACTACTCGACATCACT 2603
QY 245 SerSerLeuGlyGlnAlaSerLysThrIlePheValAlaGlyValAlaAspProAsnAsn 264
DB 2604 AGCGACAGCAGCTCGCACACCCGCTCGTGGCTGCTGCGAGTGTGACCGTGCAGCGGCACC--- 2660
QY 265 ProValPheTrp-----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAla 282
DB 2661 -----TGGTTCGACAGCCGCTGACGGGAGTACTGTGATCCGGAACCTGCGTGA 2711
QY 283 ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrlle 302
DB 2712 CCGGTGCGGTTTCCACCCCGCGCTGACCGGAGTACTGTGATCCGGAACCTGCGTGA 2771
QY 303 AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSer 322
DB 2772 GAGTTCAGCGCCAGCCCGGTTGTCAGGGGATGGACGAGTGTGTCACGGTTGCC 2831
QY 323 Val-----ThrSerGlyThrTrpTrpThrArgIleSerProValProSerThrAspThr 339
DB 2832 ACGTGTGCTGTCGACAGCGGACGCCACCGGATG-----CTCACCGCCCTG 2879
QY 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359
DB 2880 GCACAGGCTAT-----GTCCACGGCGTCCACGTCGAC----- 2912
QY 360 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 379
DB 2913 -----TGGCCCGCATCTCTC----- 2927
QY 380 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 399
DB 2928 GGCACCAACACACCGGCTACTGACCTTCGACCTAGCCTTCCACACACCGGCTAC 2987
QY 400 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 419
DB 2988 -----TGGCTC-----GAGTGGGCTCCCGCGGC 3011
QY 420 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439
DB 3012 ACGGCC-----GACTCGGGC 3026
QY 440 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 459
DB 3027 CACCCCGCTCGGACCGGAGTGGCGCTGCGCGG----- 3062
QY 460 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla---Val 478
DB 3063 TCGCGGCGCGGTTTCACGGGTTCGCGCGCGGTCGCGGACCGCGGTGTTCATC 3122
QY 479 AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 498
DB 3123 GCCGAATGGCGCTCGCCCGCGCGGACGCAC-----GACTCGCGC 3164
QY 499 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 518
DB 3165 ACGTTCGAACAGCTGACGTCACTCCGTGCGCGGATCGCGCGCGGCGG----- 3221
QY 519 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 538
DB 3222 ACGCGCAGACCTGGGTGAT----- 3242
QY 539 SerPheAspProSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGly 558
DB 3243 -----GAACCGCGCGCGGCGCGGCTTCACCGTTCACCGCGCGCGGCGG 3296
QY 559 LysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAla 578
DB 3297 GCCCGGTGACGCTGCACGCGGAG-----GGGGTTCTCCGCGCGCGCGGCTG---CCC 3347

QY 579 SerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyr 598
DB 3348 CAGCCGGAAGCGCTGCACACCGCTGGCCCGC-----CCGGGC-----GCGGTCCC 3395
QY 599 AlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 618
DB 3396 CGCGACGGCTGCTCCCGGGCGTGGCGACCGCGGACCGAGTCTTCTCGAAGCCGAAGTC 3455
QY 619 ArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyThr 635
DB 3456 GACAGC-----CCTCAGCGCTTCGTGGCACACCCGACCTGTCGACGCGGTC 3503
QY 636 PheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 655
DB 3504 TTCTCGCGGTGGCGACGGG-----AGCGCGCAGCGGACCGGATG-GCGCGACCTCGC 3556
QY 656 SerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 675
DB 3557 GGTGCACGCTGCGACCGCACCGTGTGCGCGCTGCTCACCAGCGCGACAGTGGTGT 3616
QY 676 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThr 695
DB 3617 CGTGAGCTGCGCGCTTCGACGCTCGCGAATGCCGT-----GCTCACC 3664
QY 696 GlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAlaProGlySerSer 713
DB 3665 GGAGTGGTGGACCTGGGCGAGTCCGTCGGCAGCGGATCCGACGAGTCGACGCTCT 3724
QY 714 TyrProAlaValPheValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAs 733
DB 3725 GCTTGGCTTGAAGTGTGCGGTGGCGAGGCCACCTACGACGGTGC-----GA 3775
QY 733 pAspCysGlyThrThrTrpValLeuIleAsn-----AspAspGlnHisGlnTyr 749
DB 3776 CGAGTGGCGCGGCTACACCTCATCCCGGCACACACCGGACCGCGGACCGC 3835
QY 749 rGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal----- 766
DB 3836 CACCAACCCCGCACACACACCGACCGGACCGGACCGGCTCTCACC 3895
QY 767 -----TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle----- 780
DB 3896 CCTCCAAACACCTCATCACCACCAACACACCTCATCTGTCACACACCGACCGC 3955
QY 781 ----GlyGlyAlaProSerGly-SerProSerPro-----SerValSerProSerAla 797
DB 3956 CCGAGCGCGCGCTCAGCGGCTCAGCGGCTCAGCGGCAAAACACCGCGCGCAT 4015
QY 797 exProSerLeuSerPro-----SerProSerProSerSerProSerPro 811
DB 4016 CCACCTCATCGAAACCCACCCACCCCGACCTCCCTCATCCCACTCACCACCT 4075
QY 811 exProSerPro-----SerProSerSerSerProSerSerSerProSerPro 825
DB 4076 CCACCAACCCCGCTAGCGCTCACCACACACCTCAGCGGCTCAGCGGCTCAGCGGCT 4135
QY 825 exProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 845
DB 4136 CACCAACCCCGCACACACACCGGACCGGACCGGACCGGCTCAGCGGCTCAGCGGCT 4195
QY 845 exProSerSer-----ProSerProSerSerSerProSerSerSerPro 860
DB 4196 CGCATCTCTCATCAGCGGCTCGGCGACCGCTCGCGGCTCCTCGCGGCTCCTCA 4255
QY 860 roSerProThrProSerSerPro 868
DB 4256 CCACCCCGCACCTACTCTCTCTCC 4281

RESULT 10

US-09-940-316B-28

; Sequence 28, Application US/09940316B

; Patent No. 6759536

; GENERAL INFORMATION:

; APPLICANT: KOSAN BIOSCIENCES, Inc.
 ; APPLICANT: REEVES, CHRISTOPHER
 ; APPLICANT: CHU, DANIEL
 ; APPLICANT: KHOSLA, CHAITAN
 ; APPLICANT: SANTI, DANIEL
 ; APPLICANT: WU, KAI
 ; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkdA GENE OF THE FK-520 POLYKETIDE SYNTHASE
 ; FILE REFERENCE: 30062-20026-11
 ; CURRENT APPLICATION NUMBER: US/09/940,316B
 ; PRIORITY FILING DATE: 2001-08-27
 ; PRIOR APPLICATION NUMBER: 09/410,551
 ; PRIOR FILING DATE: 1999-10-01
 ; PRIOR APPLICATION NUMBER: US 60/139,650
 ; PRIOR FILING DATE: 1999-06-17
 ; PRIOR APPLICATION NUMBER: US 60/123,810
 ; PRIOR FILING DATE: 1999-03-11
 ; PRIOR APPLICATION NUMBER: US 60/102,748
 ; PRIOR FILING DATE: 1998-10-02
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 28
 ; LENGTH: 4767
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
 ; OTHER INFORMATION: PKS synthase fragment
 ; NAME/KEY: CDS
 ; LOCATION: (3)..(4766)
 ; US-09-940-316B-28

Alignment Scores:
 Pred. No.: 3,11e-06 Length: 4767
 Score: 296.50 Matches: 233
 Percent Similarity: 35.16% Conservative: 101
 Best Local Similarity: 24.53% Mismatches: 397
 Query Match: 5.77% Indels: 223
 DB: 4 Gaps: 46

US-09-917-376-1 (1-957) x US-09-940-316B-28 (1-4767)

QY 9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28
 DB 1827 CTGACCGAACACGAGACCGCGTGGCGCCCTACCTGGCGCGGTGCGCGCGCGGATATA 1886
 QY 29 AlaAlaAlaLeuGlyValLeuProileAlaThrAlaSerProAlaHisAlaAlaThr 48
 DB 1887 CGGCGTGGCATCGACCGTG-----CGGTGACACCGTGGT-GTTCGAGCACCAGCGC 1939
 QY 49 ThrGlnProTyr----- 52
 DB 1940 CGTACTCTGGAGATGACCGTCACCGGACCGCGGTGACCGACCGACCGATCGTGT 1999
 QY 53 -----ThrTyrSerAsnValAlaIleGlyGly-----PheVal 65
 DB 2000 TGCTCTTCCCGGACGGGTGGCAGTGGTGGGATGGCGAGTGCACCTGCGCGATTCGTC 2059
 QY 66 AspGlyIleValPheAsnGly-----AlaProGlyIleLeuTyrValArgThrAsp 83
 DB 2060 GGTGTGTTCGCGGACCGATGGCGAGTGGCGGCGGTGGCGGAGTTCGT---CGA 2116
 QY 84 IleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrPheLeuLeuAspTyr 103
 DB 2117 CTGGGATCTGTTACCGTTCTGGATGATCCGCGGTGG-TGGACCGGTTCATG---TGG 2172
 QY 104 ValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123
 DB 2173 TCCAGCCC-GCTTCCTGGCGCATG-----ATGGTTTCCTGGCGCG- 2213
 QY 124 ThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSer---TyrAspProAsnAsp 142

Db 2214 -----GTGTGGCAGCGCGCGTGTGGCGCGGATGCGGTGATCGGCATTGCGCAG 2264
 QY GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln----- 156
 Db 2265 GGTGAGATCGCGCAGCTTGTGTGGCGGTGCGGTGTCTACTACGCGATCCGCCCGCGATC 2324
 QY 157 IleThrProLeuProPheLeuLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176
 Db 2325 GTGACCTTGGCAGCGCAGCGATCCCGCGGCGCTGGCGGCGCGGCGCGATGCGCATCC 2384
 QY 177 LeuAlaValAspProAsnAsnAsnIleLeuTyrPheGlyAlaProSerGlyGlyGly 196
 Db 2385 GTCGCCCTG---CCGCGCAGCATGTCAGCTGGTCGACGGGCGC----- 2426
 QY 197 LeuTyrArgSerThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProAspVal 216
 Db 2427 ---TGGATCGCGCGCCACACCGGCGCGCTCCACCGTATCGCGGCGACCCCGGAGCG 2483
 QY 217 GlyThrTyrIleAlaAsnProThrAspThrThrGly----- 228
 Db 2484 GTCGACCATGTCCTCACCGCTCATGAGGCACAAAGGGTGGCGGTGGCGGATCACCGTC 2543
 QY 229 ---TyrGlnSerAspIleGlnGlyValValTrpVal-----AlaPheAspLysSer 244
 Db 2544 GACTATGCTCGCACACCCCGCACGTCGAGCTGATCCGCGACGAATCTACTCGACATCACT 2603
 QY 245 SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264
 Db 2604 AGCGACAGCAGTCGACAGACCCCGCTCGTGGCGGTGGCGGTGGCGGTGGCGGACCC 2660
 QY 265 ProValPheTyr-----SerArgAspGlyAlaThrTrpGlnAlaValProGlyAla 282
 Db 2661 -----TGGTGCAGACCGCGCTGGACGGGAGTACTGTCACGGAACCTGCGTGAA 2711
 QY 283 ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle 302
 Db 2712 CCGGTGCGGTTTCCACCCCGCTGACCGAGTTGCGAGCGCGCGCGGCGACACCGTGTTCGTC 2771
 QY 303 AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSer 322
 Db 2772 GAGGTCAGCGCCCGCGGTGTGTGGAGCGATGGAGCATGTCCTCAGGTGGCC 2831
 QY 323 Val-----ThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThr 339
 Db 2832 ACGTGGCTGTGACGACCGCGACCGCCACCGCGATG-----CTCACCGCGCTG 2879
 QY 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359
 Db 2880 GCACAGGCTAT-----GTCCACGCGGTACCGTCGAC----- 2912
 QY 360 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 379
 Db 2913 -----TGGCCCGCCATCCCTC----- 2927
 QY 380 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 399
 Db 2928 GGCACACACACACCGCGGTACTGACCTTCGACCTAGCCCTCCACACCGCGGTAC 2987
 QY 400 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 419
 Db 2988 -----TGGCTC-----GAGTCGCGTCCCGCGCC 3011
 QY 420 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439
 Db 3012 ACGGCC-----GACTCGGCC 3026
 QY 440 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 459
 Db 3027 CACCCCGCTCCGCGACCGAGTCGCGCGCGG----- 3062
 QY 460 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla-----Val 478
 Db 3063 TCGCGGCGCGGTGTTCACGGGTCCCGGTCCCGCGGTGGCGGACCGCGGTGTTTCATC 3122

QY 811 erProSerPro-----SerProSerSerSerProSerSerSerProS 825
Db 4076 CCACCAACCCACCTACGCTCACCACCAACACACCTCCACACCCACCTACCCCAT 4135
QY 825 erProSerProSerProSerProSerProSerProSerProSerProS 845
Db 4136 CACCAACCCACCAACACCAACCAACCCACCCACCCACCTACCCCAT 4195
QY 845 erProSerSer-----ProSerProSerSerProSerSerSerP 860
Db 4196 CCGCATCTCTATCACCAGGGGCTCGGCACCTCGCGGCATCTCTCGCGCACCTCAA 4255
QY 860 roSerProThrProSerSerPro 868
Db 4256 CCACCCACACCTACCTCTCTCTCC 4281
RESULT 11
US-09-410-551B-32
; Sequence 32, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTU, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 4818
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4817)
US-09-410-551B-32
Alignment Scores:
Pred. No.: 4,76e-06 Length: 4818
Score: 293.00 Matches: 235
Percent Similarity: 33.07% Conservative: 97
Best Local Similarity: 23.41% Mismatches: 379
Query Match: 5.71% Indels: 295
DB: 4 Gaps: 48
US-09-917-376-1 (1-957) x US-09-410-551B-32 (1-4818)
QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSer----- 13
Db 1798 CTGACCGCTCCCGCGCGCGCGCTCAGCAGCCGCGGCGAAGACCTTCCTCGCTCGTGT 1857
QY 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33
Db 1858 CGGCGGTTCCCGCGGAGGCACTCGAGCAGATCGGCGCTCGCGC----- 1905
QY 34 ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrGlnProThr 53
Db 1906 -----CCTATC-----TCGACACCGCGCGCGCGCTCGACCGCGCGCGTGGCGC 1950

QY 479 AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 498
Db 3123 GCCGAACTGGCGCTCGCGCGCGCGACCGAC-----GACTGCGCC 3164
QY 499 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 518
Db 3165 ACGTGTGACAGCTGAGTCACTCTCGTCCGCGCGGATCGCGCGCGGAGGCGC--- 3221
QY 519 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleValValArgAlaGly 538
Db 3222 ACCGCGCAGACCTGGGTGAT----- 3242
QY 539 SerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGly 558
Db 3243 -----GAACCGCGCGCGCGCGCGCGCTTCACCGTCCACACCGCGTGGCGAC 3296
QY 559 LysAsnTrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAla 578
Db 3297 GCCCGTGGACGCTGCACCGCGAG-----GGGGTCTCCGCGCGCGCGGTG---CCC 3347
QY 579 SerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyr 598
Db 3348 CAGCCCGAAGCGCTGCACACCGCTGCGCCCG-----CGGCGC-----CGGTGCGC 3395
QY 599 AlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 618
Db 3396 GCGGACGGCTGCGCGCGCGTGGCGCGCGCGACCGAGTCTCGTTCGAAGCCGAAGTC 3455
QY 619 ArgSerAspArgValAsnProIlePheTyrAla-----LeuSerAsnGlyThr 635
Db 3456 GACAGC-----CCTGACCGCTTCGTGGCACACCCCGACCTGCTCGACCGGTGTC 3503
QY 636 PheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 655
Db 3504 TTCTCCGGTGGCGACCGG-----AGCCCGACCGCGCGGATG-CGCGACCTCGC 3556
QY 656 SerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 675
Db 3557 GGTGACCGCTGGGACCGCGCGTGTGCGCGCTGCTCACCCTCGCGCGCGACAGTGGTGT 3616
QY 676 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThr 695
Db 3617 CGTGGAGCTCGCGCTTCGACCGGTCCGGAATGCGGT-----GCTCACCGC 3664
QY 696 GlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAlaProGlySerSer 713
Db 3665 GGAGTCGTGACGCTGGCGGAGTGGCTCGGCGCGGATCGGACGAGTCGACGGTCT 3724
QY 714 TyrProAlaValPheValValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAs 733
Db 3725 GCTTCGGCTTGAGTGTGCTGCGGTGGGAGGCCCTACGACGCGTGC-----GA 3775
QY 733 pAspCysGlyThrThrTrpValLeuIleAsn-----AspAspGlnHisGlnTyr 749
Db 3776 CGAGTGTCCCGGAGGTATACCTCTATCATCCCGCACACACCCCGCGACCGCGACCC 3835
QY 749 rGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgVal----- 766
Db 3836 CACCAACCCCAACACACACCCACCGACCCACCAACCAACACACCGGTCTCTACCGC 3895
QY 767 -----TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle----- 780
Db 3896 CCTCCACACACCTCATCACCAACACACACCTCATCGTCCACACACCGACCGACCC 3955
QY 781 ----GlyGlyAlaProSerGly-SerProSerPro-----SerValSerProSerAlaS 797
Db 3956 CCGAGCGCGCGGTCAACCGGCTCACCAGCGACCGCACAAACAAACACCGCGCGCAT 4015
QY 797 erProSerLeuSerPro-----SerProSerProSerSerProS 811
Db 4016 CCACCTCATCGAAACCCACACACCCCGACCGCGCGCGCTCACCCTACCACTACCACT 4075

QY 54 -----TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71
 Db 1951 AGACACTGG----- 1959
 QY 72 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyr----- 88
 Db 1960 -----CCCGGC-----GTACGCACCTTACCCACCGGGCGTACTGTCTCGGG 2001
 QY 88 ----- 88
 Db 2002 ACACCGTCATCGCGCTCCCCCGGAGCAGCGCGAGCAACTGTCTTCTACTCCG 2061
 QY 89 -----ArgTTPAspAla 92
 Db 2062 GTCAGGGCACCAGATCCCGCGATGGCGGAGAGCTAGTGCATGCTCGGTGGTTCG 2121
 QY 93 AlaAsnGly----- 95
 Db 2122 CCGAGCGGATGGCCGAGTGTGGCGGGTTCGCGAGTTCGTGGACTGGGATCTGTTCA 2181
 QY 96 -----ArgTTPleProLeuLeuAspTTPValGlyTTPAsnAsnTTP 109
 Db 2182 CGGTTCTGGATGATCCCGCGGTGGTGGACCGGGTTGATGTGTCCAGCCC-GCTTCTCTGG 2240
 QY 110 GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTTPAla 129
 Db 2241 GCGATG-----ATGTTTCTCTGCGCG-----GTGTGGCAG 2273
 QY 130 AlaValGlyMetTyrThrAsnSer---TTPAspProAsnAspGlyAlaIleLeuArgSer 148
 Db 2274 CGGCGGTGTGGCGCGGATGGTGTATCGGCATTCGCGGATGAGATCGCGCAGCT 2333
 QY 149 SerAspGlnGlyAlaThrTTPGln-----IleThrProLeuProPhe 162
 Db 2334 TGTGTGGCGGTGGTGTCTACTAGCGATGCGCGCGGATGTCGACCTTGGCGCAGCC 2393
 QY 163 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn 182
 Db 2394 CGCATCGCCCGCGGCGTGGCGCGCGGCGGATGGATGCGTGGCGCTG-----CCCGCG 2450
 QY 183 AsnAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTTPArgSerThrAsp 202
 Db 2451 CAGGATGTGAGTGTGTGACGGGCGC-----TGGATCGCCGCCAC 2492
 QY 203 SerGlyAlaThrTTPSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn 222
 Db 2493 AAGCGCGCGCTCCACCGTGTATCGCGGCGCACCCCGAAGCGGTGACCATGTCTCTACC 2552
 QY 223 ProThrAspThrGly-----TyrGlnSerAspIle 233
 Db 2553 GCTCATGAGGCACAAAGGGGTGGGGTGGCGGATCACCGTTCGACTATGCTCGCACACC 2612
 QY 234 GlnGlyValValTTPVal-----AlaPheAspLysSerSerSerLeuGlyGln 250
 Db 2613 CGGACGTGAGTGTATCGGACGAACTACTGCAATCATTAGCAGCAGAGCTCGCAG 2672
 QY 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTTP----- 268
 Db 2673 ACCCGCTCGTGGTGTGTGCGCGTGGACCGCGCGC-----TGGGTGCGAC 2720
 QY 269 SerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPheIlePro 288
 Db 2721 AGCCCGCTGGAGGGGAGTACTGTGTACCGGAACCTCGGTGAACCGGTTCGTTTCCACCCC 2780
 QY 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308
 Db 2781 GCCGTACCGAGTTCAGGCCCGGCGGACACCGTGTTCGTGAGGTGACGCCGCGCCG 2840
 QY 309 GlyProTyrAspGlySerSerGlyAspValTTPLysPheSerVal-----ThrSer 325
 Db 2841 GTGTTGTTCAGCGCATGACACAGATGTCTGTACCGGTGCGCATCGTGTGTGACGAC 2900

QY 326 GlyThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 345
 Db 2901 GCGCAGGCCACCCGGATG-----CTCACCGCCTGCACAGGCTAT----- 2942
 QY 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365
 Db 2943 GTCCACGGCGTACCGCTGCAC----- 2963
 QY 366 SerTTPProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArg 385
 Db 2964 -----TGGCCCGCATCTC-----GGCACCCACACACCCCG 2996
 QY 386 IleTTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 405
 Db 2997 GTACTGGACCTTCGACCTTCCACCTTCCACACACCGCGGTAC----- 3038
 QY 406 GluProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGly 425
 Db 3039 -----TGGCTC-----GAGTGGCTTCCCGCGCCACCGGCC----- 3068
 QY 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
 Db 3069 -----GACTCGGGCCACCCGCTCTCGGCACC 3095
 QY 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHis 465
 Db 3096 GGAGTTCGCGTGGCGG-----TCGCGCGCGCGGTGTTTC 3131
 QY 466 IleAlaProMetValLysGlyLeuGluThrAla---ValAsnAspLeuIleSerPro 484
 Db 3132 ACGGTCGCGTGGCGCGCGTGGCGCGGTTCATCGCCGAACTGGCGCTCGCC 3191
 QY 485 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp 504
 Db 3192 GCGCGCGACGCCAC-----GACTCGCCACCGTTCGACAGCTCGAC 3233
 QY 505 ValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524
 Db 3234 GTCACTCGTGGCGCGGATCCCGCGCGCGAGGCC---ACCGCGACAGCTGGGTG 3290
 QY 525 AspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSer 544
 Db 3291 GAT-----GAACCGCGCGC 3305
 QY 545 GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTTPPheGlnGly 564
 Db 3306 GACGGCGCGCGCTTACCGTCCACCCCGTGGCGAGCCCGCGTGGAGCTGCGAC 3365
 QY 565 SerGluProGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArg 584
 Db 3366 GCGGAG-----GGGTTCTCGCGCGCGCGGTG---CCCCAGCCGAGCGCTCGAC 3416
 QY 585 PheValTTPAlaProGlyAspProGlnProValValTyrAlaValGlyPheGlyAsn 604
 Db 3417 ACCGCTGCGCGCGC-----CGGCGC-----GCGGTGCGCGCGAGCGGTGCGCGG 3464
 QY 605 SerTTPAlaAspSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 624
 Db 3465 GCGTGGCGACGCGCGACAGGTCTTCTGTCGAAGCCGACGTCGACAGC----- 3512
 QY 625 ProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAsp 641
 Db 3513 CTGAGCGGTTCGTGGCACACCCCGACCTGCTCGACGCGGTCTTCTCGCGGTGCGCGC 3572
 QY 642 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 661
 Db 3573 GGG-----AGCGCGCGCGCGCGGATG-GCGCGACCTCGCGGTGCGACGCTCGACGC 3625
 QY 662 MetPheHisAlaValProGlyLysGlyAspLeuTTPLeuAlaAlaSerSerGlyLeu 681
 Db 3626 CACCGTGTCTGCGCGCTGCTCACCGCGCGCGACAGTGTGTGTGTGAGCTCGCGCTT 3685
 QY 682 TyrHisSerThrAsnGlySerSerTTPSerAlaIleThrGlyValSerSerAla--- 700

3686 CGAGGTCCGGATGCCGT-----GCTACCGCGAGTCTGGTACGCTGG 3733
701 ---ValAsnValGlyPheGlySerAlaProGlySerTyrProAlaValPheVal 719
3734 CGAGTCCGTCCGCGAGCGATCCGAGAGTCCGAGCGTCTGCTTCGGCTTGAGTGGTT 3793
720 ValGlyThrIleGlyVal-ThrGlyAlaTyrArgSerAspCysGlyThrThr 739
3794 GCGGTGGCGGAGCCACTAGCGGTGCC-----GACGAGCTGCCGAGGGCTA 3844
739 pValLeuIleAsn-----AspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAl 755
3845 CACCCTCATACCGCCACACACACCCCGAGACCCCGAGACCCCAACCCCAACAC 3904
755 alleThrGlyAspHisAlaAsnLeuArgVal-----TyrIle 768
3905 ACCACACGACCCACACACACAAACACAGCGTCTCAGCGCCCTCCAAACACCATCAT 3964
768 eGlyThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaProSe 785
3965 CACCACCAACACACCTCATGCTCCACACACACACACCCCGAGCGCGCGTCCAC 4024
785 rGly-SerProSerPro-----SerValSerProSerAlaSerProSerLeuSerPro- 802
4025 CGGCTCACCAGCCACACCGACAAACGACACCCCGCGCGCATCCACCTCATCGAAACCA 4084
803 -----SerProSerProSerSerProSerProSerPro----- 814
4085 CCACCCCAACACCCACTCCCTCACCACCTCACCACCTCCACCAACCCCACTAGC 4144
815 -----SerProSerSerProSerSerProSerProSerProSerProSerProS 831
4145 COTCACCACAAACACCTCCACACCCCGACCTCACCACCCCATCCACCAACCAACAC 4204
831 erProSerProSerArgSerProSerProSerProSerProSerProSerProS 849
4205 CACCACAAACCCCAACACCCCGACCTCCACCCCAACCCCAACCCCAACCAACCG 4264
850 -----ProSerProSerSerProSerProSerProSerProSerProSerProS 866
4265 CGCTCCGCGACCTCCGCGGATCTCTCGCCGCGACCTCCTACCAACCCCAACCAACCT 4324
866 erSerPro 868
4325 CCTCTCCC 4332

RESULT 12
US-09-940-316B-32
; Sequence 32, Application US/09940316B
; Patent No. 6759536
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940,316B
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/410,551
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 32
; LENGTH: 4818
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKs synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4817)
US-09-940-316B-32

Alignment Scores:
Pred. No.: 4,76e-06 Length: 4818
Score: 293.00 Matches: 235
Percent Similarity: 33.07% Conservative: 97
Best Local Similarity: 23.41% Mismatches: 379
Query Match: 5.71% Indels: 295
DB: 4 Gaps: 48

US-09-917-376-1 (1-957) x US-09-940-316B-32 (1-4818)

QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSer----- 13
Db 1798 CTGGACCGCTCCCGCGCGCGCGCTCAGACACCGCGGCGAAGACCTTCCGCTGCTCGTGT 1857
QY 14 ArgArgLeuValSerLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33
Db 1858 CGCGCGGTTCCTCCCGGAGGCATCGACGACGATCGGCGCGCTGCGCG----- 1905
QY 34 ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThr 53
Db 1906 -----CCTATC-----TCGACACCGCGCGCGCTCGACCGCGCGCGCTGCGCG 1950
QY 54 -----TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71
Db 1951 AGACACTGG----- 1959
QY 72 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyr----- 88
Db 1960 -----CCGGC-----GTACGCACTTCAACCCACCGCGCGGTACTGCTCGGG 2001
QY 88 ----- 88
Db 2002 ACACCGTCATCGCGCTCCCGCGCGGACGAGCCGACGAACTCGTCTTCTACTCTCCG 2061
QY 89 -----ArgTrpAspAla 92
Db 2062 GTCAGGCGCACCGACATCCCGCGATGGCGGCGAGCTAGCGGATTCGTCGGTGGTTCG 2121
QY 93 AlaAsnGly----- 95
Db 2122 CCGAGCGGATGGCGGAGTGTGCGCGCGCTTGGCGAGTTCGTGGACTGGGATCTGTTC 2181
QY 96 -----ArgTrpIleProLeuLeuAspTrpValGlyTyrAsnAsnTrp 109
Db 2182 CGGTCTCGATGATCCGCGGCTGTGGACCGGGTGTATGTGGTCCAGCCC-GCTTCTCG 2240
QY 110 GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAla 129
Db 2241 GCGATG-----ATGGTTTCCCTGCGCG-----GTGTGGCGAG 2273
QY 130 AlaValGlyMetTyrThrAsnSer-----TrpAspProAsnAspGlyAlaIleLeuArgSer 148
Db 2274 GCGGCGGTGTGGCGCGGATGCGGTATCGGCCATTCGCGAGGTGAGATCGCGCGAGCT 2333
QY 149 SerAspGlnGlyAlaThrTrpGln-----IleThrProLeuProPhe 162
Db 2334 TGTGTGGCGGTGCGGTGTCTACATCGCGATGCGCGCGCGGATCGGTGACCTTGGCGAGCCAG 2393
QY 163 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn 182
Db 2394 GCGATCGCGCGGCTGCGCGCGCGCGCGCGATGCGCATCGCTCGCGCGCTG---CCCGCG 2450


```
QY 183 AsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAsp 202
DB 2451 CAGGATGTCGAGCTGGTCGAGGGGCC-----TGGATCGCCGCCAC 2492
QY 203 SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn 222
DB 2493 AACGGGCGCGCTCCACCGTATCGCGGACCCCGGAGCGGTGCACCATCTCTCACC 2552
QY 223 ProThrAspThrThrGly-----TyrGlnSerAspIle 233
DB 2553 GCTCATGAGGCACAAGGGGTGGGTGGCGGATCACCGTCGACTATGCTCGCACACC 2612
QY 234 GlnGlyValValTrpVal-----AlaPheAspLysSerSerSerLeuGlyGln 250
DB 2613 CCGCAGCTCGAGCTGATCCGCGAGAACTACTCGACATCACTAGCAGCAGTCTCGCAG 2672
QY 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp----- 268
DB 2673 ACCCGCTGCTGCGTGGCTGACCGTGCAGCGGCACC-----TGGGTGCAC 2720
QY 269 SerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro 288
DB 2721 AGCGCGTGGACGGAGTACTGTGTACCGGAACCTGCGTGAACCGGTGCGTTTCCACGCC 2780
QY 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308
DB 2781 GCCGTCAGCCAGTTGACAGCCCGAGCGGCACACCGTGTTCGTGCAGAGTCAGCGCCACCG 2840
QY 309 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal-----ThrSer 325
DB 2841 GTGTTGTGACGGATGAGACAGATGCTCAGCGTGGCCAGCTGCGTGTGCAGC 2900
QY 326 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 345
DB 2901 GCGCAGCGCCACCGCATG-----CTCACCGCGCTGGCACAGCGCTAT----- 2942
QY 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValaThrGlnIle 365
DB 2943 GTCCACGCGCTCACCGTCGAC----- 2963
QY 366 SerTrpTrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 385
DB 2964 -----TGCGCCGCCATCCTC-----GGCACCCACACACCGG 2996
QY 386 IleTrpAspTrpTrpSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 405
DB 2997 GTACTGGACCTCCGACCTACCGCTTCCACACACCGCGGTAC----- 3038
QY 406 GluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGly 425
DB 3039 -----TGCGTC-----GAGTCGGTCTCCCGCGCACCGGC----- 3068
QY 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
DB 3069 -----GACTCGGGCCACCCCGTCTCTCGGCACC 3095
QY 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 465
DB 3096 GGAGTCGCGGTGCGCGG-----TGCGCGGCGCGGTGTC 3131
QY 466 IleAlaProMetValLysGlyLeuGluThrAla-----ValAsnAspLeuIleSerPro 484
DB 3132 ACGGGTCCCGTCCCGCGGTGCGGACCGCGCGGTTCATCGCGGAATCGCGCTCGCC 3191
QY 485 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp 504
DB 3192 CCGCGCGACGCCACC-----GACTCGCGCACCGGTTCGAACAGCTCGAC 3233
QY 505 ValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerVal 524
DB 3234 GTCACTCGGTGCGCGGATTCGCCCGCGCGGAGGGCC-----ACCGCGCAGACCTGGGTG 3290
```

```
QY 525 AspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSer 544
DB 3291 GAT-----GAACCGCGCGCC 3305
QY 545 GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly 564
DB 3306 GACGGCGCGCGCTTCCACCGTCCACACCGCGTCCGCGACCGCCCGTGGACGCTGCAC 3365
QY 565 SerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg 584
DB 3366 CCGGAG-----GGGTTCCTCGCGCCCGCGCGGTG-----CCCCAGCCCGGACCGTGCAC 3416
QY 585 PheValTrpAlaProGlyAspProGlnProValValTyrAlaValGlyPheGlyAsn 604
DB 3417 ACCGCTGGCCCGC-----CGGGC-----GCGGTGCCCGCGGCGGTGCGCGG 3464
QY 605 SerTrpAlaAspSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 624
DB 3465 GCGTGGCGACGCGCGGACCGAGTCTTCGTGGAAGCGGAAGTCGACAGC----- 3512
QY 625 ProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAsp 641
DB 3513 CTTGACGCGCTTCTGTCGACACCGCGACCTGCTCGACGCGGTCTTCTCCGCGTTCGCGCAG 3572
QY 642 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 661
DB 3573 GGG-----AGCGCGCAGCGGACCGGATG-GCGCGACCTCGCGGTGCACCGTTCGCGCAGC 3625
QY 662 MetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAspSerSerGlyLeu 681
DB 3626 CACCGTGTGCGCGCTCCTCACCGCGCGACAGTGTGTGTCGAGCTCGCGCGCTT 3685
QY 682 TyrHisSerThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSerAla--- 700
DB 3686 CGACGCTCGCGAATGCGCGT-----GCTCACCGCGGAGTCGGTGACGCTGGG 3733
QY 701 ---ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719
DB 3734 CGAGTCCGTCGCGGCGGATCCGACGAGTCCGACGCTGCTTCGCTTCGCTGAGTGGTT 3793
QY 720 ValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAspCysGlyThrThrTr 739
DB 3794 GCGCGTGGCGGAGGCCACTACGCGGTGCC-----GACGAGCTCGCGGAGGCTA 3844
QY 739 pValleuIleAsn-----AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAl 755
DB 3845 CACCTCATACCGCGCACACACCGCGACCGCGACCGCGACCGCGCGCGCGCGCGCG 3904
QY 755 aIleThrGlyAspHisAlaAsnLeuArgVal-----TyrIle 768
DB 3905 ACCCACGCGACCCACACACACACACCGCGTCTCTACCGCGCTTCCACACCGCGCTCAT 3964
QY 768 eGlyThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaProse 785
DB 3965 CACACCAACACACACCTCATCTCCACACCGCGCGCGCGCGCGCGCGCGCGCG 4024
QY 785 rGly-SerSerPro-----SerValSerProSerSerAlaSerProSerLeuSerPro- 802
DB 4025 GCGCGCTCACCGCGACCGCGCAAAAGAAACACCGCGCGCGCGCGCGCGCGCGCG 4084
QY 803 -----SerProSerProSerSerSerProSerProSerPro- 814
DB 4085 CCACCCCGCACCGCGCGCTTCCCGCTCACCGCACTCCACCGCGCGCGCGCGCGCGCG 4144
QY 815 -----SerSerSerSerSerSerSerSerSerProSerProSerProS 831
DB 4145 CCTCACCACACACCTTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4204
QY 831 erProSerProSerArgSerProSerProSerAlaSerProSerProSerSerSer- 849
DB 4205 CACACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4264
QY 850 -----ProSerProSerSerSerSerSerProSerProSerProSerS 866
```


Db 4265 CGGTCGGCACCCTCGCGGATCTCGCCGCCACCTCAACACCCCACTACT 4324
 QY 866 erSerPro 868
 Db 4325 CCTCTCCC 4332

RESULT 13

US-09-252-991A-13773/c
 ; Sequence 13773, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIORITY FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 13773
 ; LENGTH: 4131
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-13773

Alignment Scores:
 Pred. No.: 6,77e-06 Length: 4131
 Score: 288.50 Matches: 246
 Percent Similarity: 31.64% Conservative: 109
 Best Local Similarity: 21.93% Mismatches: 376
 Query Match: 5.62% Indels: 392
 DB: 4 Gaps: 56

US-09-917-376-1 (1-957) x US-09-252-991A-13773 (1-4131)

QY 37 IleAlaIleThrAlaSerProAlaHisAlaAlaThrGlnProTyrThrTrpSer--- 55
 Db 4113 GTCACGTACCGGACGACGCGCGCGGCAATACACGCGTCCCGCTACCACGCGTG 4054
 QY 56 -----AsnValAlaIleGlyGlyGlyGlyPheValAspGly 67
 Db 4053 GATTCCTCGTCGTCGATCCGCGAGTGGATCCGAGCAACGGTTTCGGTGATCAGCGC 3994
 QY 68 IleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMet 87
 Db 3993 -----ACCGCGACGCGCGCAACACC 3973
 QY 88 TyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsn 107
 Db 3972 ATCATCATCAGATGGCAACGCAACCCGATTTGGCCAGGTCAACCGCGACGCGGT 3913
 QY 108 AsnTrpGlyTyr-----AsnGlyValVal-----SerIleAla 118
 Db 3912 AACTGGTCTTCACTCCAGGCATCCCGTCGCGGATGGACGCGTGTCAACGCTGGTGGC 3853
 QY 119 AlaAspProIleAsnThrAsnIysValTrpAla-----AlaValGlyMet 133
 Db 3852 CGCAGCCCAAGCAATGTCACAGATGCGCGCGCGGTGATCACTGTGTGATGCGTGGCCCG 3793
 QY 134 TyrThrAsnSerTrpAspProAsnAspGlyAla---IleLeuArgSerSerAspGlnGly 152
 Db 3792 GCGGCGCGGTGATCATCGACCAACGCGCACCGAGATAGCGGTACCGCGGAGCGCGC 3733
 QY 153 AlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGly 172
 Db 3732 GCGACGGTATCTCACCAGT-----GGCGGCGCAACCG----- 3697
 QY 173 MetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaPro 192

Db 3696 ATCGGCCAGGCCACCGCC----- 3679
 QY 193 SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsn 212
 Db 3678 GACGGCAGCGCAACTGGACGTTCACTCCG----- 3649
 QY 213 PheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAsp 232
 Db 3648 -----GGCACCCTGCGTGGCCAAACGGCACCGTGATCAACGCC----- 3613
 QY 233 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSer 252
 Db 3612 -----GTGGCCACAGGACCCCGCGCAATACCAACGCGTCCGCGCAGC 3571
 QY 253 LysThrIlePhe-----ValGlyValAlaAspProAsnAsnProValPhe 267
 Db 3570 GTCACCGTGCATGCCATGCCCGCGCGCGGTGATCAATCCGACGACGAGATCGTC 3511
 QY 268 TrpSer----- 269
 Db 3510 ATCAGCGGTACGGCGGAACCCGGGGCCACGGTGATCTCACCGACGCAACGCCACCCG 3451
 QY 270 -----ArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro--- 283
 Db 3450 ATCGGCCAGGTCAACCGCGACGCGCAGCGCAAGTGGGCTTTCAGCGCGCCACCGCGT 3391
 QY 283 ----- 283
 Db 3390 GCCAATGGCAGCGTGATCAATGCGTGGCCCGACGACGCGCGCGCAACACAGCAGTCCC 3331
 QY 284 -----ThrGlyPheIleProHisGlyValPheAspProValAsnHis 298
 Db 3330 ACCAGCGCCACGTCGATCGCTGGCGCCAGACGCGCGGTGATCGATCGACCAACGGT 3271
 QY 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySerSer 315
 Db 3270 AGCGTGATCGCGGTACCGCGAGCGGTGTGTGCGCACCGTGATCTCACCGACGCAACGCC 3211
 QY 316 GlyAspValTrpLysPheSerValThr---SerGlyThrTrpThrArgIleSerProVal 334
 Db 3210 AACCCGATCGCGCAGGTCAACCGCGATGGCAGCGCAACTGGAGCTTCACCGCGCGCACG 3151
 QY 335 Pro-----SerThrAspThrAlaAsnAspTyrPhe 344
 Db 3150 CGCTCTCCATGGCAGCGTGTCATGCGGTGGCCGCGCGCGCGGTGATCGACCCGACACC 3091
 QY 345 GlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 356
 Db 3090 GGCCCGCGCAGCACCGCGTGGACTCGGTGGCGCGCGCGCGGTGATCGACCCGACG 3031
 QY 357 AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSer 376
 Db 3030 AACCGCAGCGTGATCGCGGTACCGCGAAGCGGTGCGACGCGTGTGATCTCTC----- 2980
 QY 377 ThrAspGlyGly-----AlaThrTrpThrArgIleTrpAspTrpThr 390
 Db 2979 ACCGATGGCGCGGCAACCCGATCGCGCAGCGCACCGCGCGATGGCAGCGCACTGGAGC 2920
 QY 391 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla----- 405
 Db 2919 TTCACCCCGCGCAGCGCGTGGCCAAACCGCACCGGTGATCAATGCGGTGGCCAGGATCCG 2860
 QY 406 -----GluProTrpLeuThrPhe-----GlyValGlnProAsnProPro 418
 Db 2859 GCGCGCAATACCAAGCGCGCCAGCACCGACCGTGGACGCGGTGGCGCGCGCCGCCCGC 2800
 QY 419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 438
 Db 2799 TGGTCAACCCGAGCAACCGC----- 2779
 QY 439 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 455
 Db 2778 ---AGCGTGATCGCGGTACCGCGGAGCGCGCGCGCGGTGATCTCTCACC----- 2731


```
Qy 456 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 475
Db 2730 -----GACGCGCGGCAAC-----CGATCGCCAGGTACACCGCCGAC 2692
Qy 476 ThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu----- 490
Db 2691 GGCAGCGGCACTGGAGTTTCACGCC-----GGCAGCGGCTGGCCACCGGTCGGTG 2638
Qy 491 IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAlaVal 508
Db 2637 ATCAATCGCTGGCCAGGACCGCGCGCAACACACAGCGCGCGCGCAGCACACCGGTG 2578
Qy 509 ProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerValAsp--- 525
Db 2577 GACTCGGTAGCCCGCGCCACCCCGGTGCTGATCCGAGCAACCGGTGCGGTACAGCGGT 2518
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545
Db 2517 ACCGCCAGAGCGGGCCACGGTGATCTCACCGAGCGGCGGCAACCCCGATACGGCAG 2458
Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 565
Db 2457 GCC-----ACGCCGATGGCAGCGCAACTGG-----AGCTTC 2425
Qy 566 GluProGlyGlyValThrThrGlyGlyThrVal-----AlaAlaSerAlaAspGlySer 583
Db 2424 ACTCCGGGCACACCGCTGACCAACCGCACCGGTGATCAATGCGGTGGCCACGAGCAGCGCC 2365
Qy 584 ArgPheValTrpAlaPro----- 589
Db 2364 GGCAACACGACCGGTCCGTCAGTGCACCAACAGTGCAGCGGTGGCCCGCCACCCCGGTG 2305
Qy 590 -----GlyAspProGlyGlnProValVal 597
Db 2304 ATCGACCCGCAATGGTGTCAAACTCAGCGGCACCGCGCAACCGCGGTCCGGTGATC 2245
Qy 598 TyrAlaValGlyPheGlyAsn-----SerTrp 606
Db 2244 CTCACCGATGGCAATGGCAACCCGATCGCCGAGACCTCGCCGCGGTAGCGGCAACTGG 2185
Qy 607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 626
Db 2184 ACCTTCACCGCGGACCGCG----- 2164
Qy 627 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 646
Db 2163 -----CTGCCAACGGCACTGGTGTCACAGCC-----GTGCCCCAG 2128
Qy 647 GlnProValAla---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 665
Db 2127 GACCGCGCGGCAATATACCAGCGGTCCGCGCACCGACCGGTGGATACGTCGCGCGCGCC 2068
Qy 666 ValProGlyLysGluGlyAspLeuTrpLeuAlaAspSerSerGlyLeuTyrHisSerThr 685
Db 2067 ACGCG-----GTATCAATGCCAGC 2047
Qy 686 AsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPhe 705
Db 2046 AACGCG-----AGCGTGATCACCGCACCGCC-----GAGGTCCG--- 2011
Qy 706 GlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGly 725
Db 2010 -----GCCAAAGTGATCCTCACCGACGGCAACCGCAAC 1978
Qy 726 ValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspAsp 745
Db 1977 CCGATCGCGCAGACACCGCGCAGCGAGTGGC---AACTGGACTCTCACCCCGCGCAGC 1921
Qy 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 765
Db 1920 CCGTGGCCAAAGTACGGTGATCAACGCCGTGCGCGAAGACCGCGCG----- 1873
```

```
Qy 766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaProSer 785
Db 1872 -----GCCAACGCCAGCGGT---CGGCGC 1852
Qy 786 GlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProSe 805
Db 1851 AGCACCGCGGTGGACTCGGTGGCGCGTCCGCT---CGCTGCTGAGCATCAGCGCGAC 1795
Qy 805 rProSerSerSerProSerPro-----SerProSerProSe 817
Db 1794 GCGCGGTGCTGACCGGCAACCGCGAGCGGCAACAGTCAAGTGGTGTCTGTTGTCACGCG 1735
Qy 817 rSerSerProSerProSerProSerProSerProSerProSerProSerProSe 837
Db 1734 GACACCGCAACCGCATCAGGTACCGTTCAGCGCGCGCGCAACTTCAGCTTCGCGTTC 1675
Qy 837 r-----ProSerPr 840
Db 1674 GCGCGCGCTGATCACCGGCAACTGATCGCGGGGTTGCGTGGACCGCGCGCAAC 1615
Qy 840 oSerAlaSerProSerProSerSerProSerProSerProSerProSerProSe 859
Db 1614 GTCAGCGCGCGCGCCACCATCAACCGCGGACCTGGCGCCCGCCCGCATCAGCGTCCG 1555
Qy 859 rProSerProThrProSerSerProValSerGlyGlyValLysValGlnTyrLysAs 879
Db 1554 GAAGCGCGCGATCTGGATCAACCGCGGAGATCGGGGACGGC-ATCCAGTTC----- 1501
Qy 879 nAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGl 899
Db 1500 -----GATGTACGCTCGTCCGACCATCAGTCCGCG----- 1468
Qy 899 ySerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGl 919
Db 1467 -----CAGGTGCTCAGCGTCAAGTTC-----GCCGGGCAAGAGCGCTA 1430
Qy 919 ySerSerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAl 939
Db 1429 CGAGCGCGAGTTCAGCATACCTCACCGCGCGGACATCGCGCGCGCAACTGACCTT 1370
Qy 939 aSerPhe-----GlySerValAsnPro-----AlaThrProThrAl 951
Db 1369 GACCTGACGCTCCCGCGCGCATGCGCGCTTCGCGAGGGCGCTCAGCGTCCACCGC 1310
Qy 951 aAsp 952
Db 1309 CGAC 1306
```

RESULT 14
US-09-410-551B-18
; Sequence 18, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18

QY 84 IleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyr 103
DB 1952 TGGGATCTGTTCACGGTCTCGATGATCCGGCGGTGG-TGACCGGTTGATG---TGG 2007
QY 104 ValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123
DB 2008 TCCAGCCC-GCTTCTCGCGCATG-----ATGGTTTCCCTGGCGCG- 2048
QY 124 ThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSer---TyrAspProAsnAsp 142
DB 2049 -----GTGTGCGACGCGCGGTGTGCGCGCATGCGGTGATCGGCATTCGCAG 2099
QY 143 GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTyrGln----- 156
DB 2100 GGTGAGATCGCGCATGTGTGCGGGTGGGTGCTCACTACGCGATGCGCGCGCATC 2159
QY 157 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176
DB 2160 GTGACCTTGGCAGCAGCAGCATCCCGGGGCTGGCGGGCGGGCGCATGCCATCC 2219
QY 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196
DB 2220 GTGCGCCCTG---CCGCGCAGCATGTGAGCTGGTCGACGGGCGC----- 2261
QY 197 LeuTyrArgSerThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProAspVal 216
DB 2262 ---TGGATCGCGCCGCACACAGCGGCGCCCTCCACCGTGATCGCGGGCACC----- 2318
QY 217 GlyThrTyrIleAlaAsnProThrAspThrThrGly----- 228
DB 2319 GTCGACCATGTCTCACCGCTCATGAGGCACAGGGGGTGGGGTGGGGGATCACCGTC 2378
QY 229 ---TyrGlnSerAspIleGlnGlyValValTyrVal-----AlaPheAspLysSer 244
DB 2379 GACTATGCTCGACACCCCGACGTGAGCTGATCCGCGAGAACTACTCGACATCACT 2438
QY 245 SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264
DB 2439 AGCGACAGCAGCTCGACACCCCGCTGTCGCGGTGGTGTGACCGGTGGACCGCACC--- 2495
QY 265 ProValPheTyr-----SerArgAspGlyGlyAlaThrTyrGlnAlaValProGlyAla 282
DB 2496 -----TGGGTGACAGCCCGCTGGACGGGGAGTACTGTACCGGAACCTCGGTGAA 2546
QY 283 ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle 302
DB 2547 CCGGTGCGTTTCACCCCGCTCAGCCAGTTGACAGGCCCGCGCGACACCGTGTGTC 2606
QY 303 AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSer 322
DB 2607 GAGGTGAGCGCCAGCCGCGGTGTGTGACGGCGATGGACGAGATGTCTCACGGTTGCC 2666
QY 323 Val-----ThrSerGlyThrTyrThrArgIleSerProValProSerThrAspThr 339
DB 2667 ACGTGTGCGTGTGACGACGGCAGCGCCACCCCGATG-----CTCACCGCCCTG 2714
QY 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359
DB 2715 GCACAGCCCTAT-----GTCCACGGCGTCACGTCGAC----- 2747
QY 360 MetValAlaThrGlnIleSerTyrProAspThrIleIlePheArgSerThrAspGly 379
DB 2748 -----TGGCCCGCATCTCTC----- 2762
QY 380 GlyAlaThrTyrThrArgIleTyrAspThrTyrSerTyrProAsnArgSerLeuArgTyr 399
DB 2763 GGCACCAACCAACCGGGTACTGGACCTTCCGACCTACGCTTCCACACCGCGCGGTAC 2822
QY 400 ValLeuAspIleSerAlaGluProTyrLeuThrPheGlyValGlnProAsnProProVal 419
DB 2822 ----- 2822

QY 420 ProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439
DB 2823 -----TGGCTCGAGTCCGACGC-----CCGGCCGATCCTCCGAC 2855
QY 440 -----ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys 457
DB 2856 CGGGCCACCCCGTGTGCGGTATCGCCCTCGCCGG----- 2897
QY 458 TrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla 477
DB 2898 -----TCCCGCGCGGGTGTTCACGGT----- 2921
QY 478 ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497
DB 2922 -----TCCGTGCGACCGGTGGCGACCGCGCGGTTCGTGCGCGAGCTG 2966
QY 498 GlyGlyPheThrHisAlaAspVal-----ThrAlaValProSerThrIlePheThrSerProVa 517
DB 2967 GCGCTGCGCGCGCGGACGCGGTGCTGCGCACCGTTCGAGCGGTTCGACATCGCCTCC 3026
QY 517 IPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValAlaArgAl 537
DB 3027 GTGCGCGCGCGCGCGGCATGCGCGACGACCGTACAGACCT----- 3069
QY 537 aGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGln 557
DB 3070 -GGGTGAC-GAGCCGGCGGACGACCGCGCGCGCTTACCCTGTCACACCGCACCGG 3127
QY 557 yGlyLysAsnTyrPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAl 577
DB 3128 CGACGCGCGTGGACGCTGCACCGCAG---GGGTGCTGCGCGCCCATGGCAGC----- 3179
QY 577 aAlaSerAlaAspGlySerArgPheValTyrAlaProGlyAspProGlyGlnProValVa 597
DB 3180 -GCCCTGCGCGATGCGCGCGACCGCGAGTGGCCCCA-----CCGGC-----GCGGT 3226
QY 597 lTyrAlaValGlyPheGlyAsnSerTyrAlaAlaSerGlnGlyValProAlaAsnAlaGl 617
DB 3227 GCGCGGAGCGGCTCGCGGTGTGTGGCGCGGGGACGAGTCTTCGCGGAGCGCA 3286
QY 617 nIleArgSerAspArg-----ValAsnPro-----LysThrPheTyrAl 630
DB 3287 GGTGGACGACCGACCGGTTTCTGTGTGTCACCCCGACCTGTCTGACGCGGTCTTCTCGC 3346
QY 630 aLeuSerAsnGlyThr----- 635
DB 3347 GGTGCGCGACGGAAGCG 3406
QY 636 -----PheTyrArgSerThrAspGlyGlyValThrPheGlnPr 648
DB 3407 CGCCACCGTACTCGCGCGCTGCTCACC CGCGCACCGAGCGCGCGCGCGCGCGCGCGCG 3466
QY 648 oVal---AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValPr 667
DB 3467 CTTTCGACGGCGCGCGCTGCGGTACTCACCGCGGAGCGGTGACGCTGCGGGAGGTGGC 3526
QY 667 oGlyLysGluGlyAsp-----LeuTrpLeuAlaAlaSe 678
DB 3527 GTCACCGTCCGCTCCGAGGAGTGGACGCGCTGCACCGGTGGAGTGGCTCGCGGTGCG 3586
QY 678 rSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTyrSerAlaIleThrGlyValSe 698
DB 3587 CGAGGGGTCTAC-----GACGGTGACCTGCGCGGAGGAGCATGTC-----CTGAT 3631
QY 698 rSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPh 718
DB 3632 CACCGCGCGCCACCCCGACGACCCCGGAGGACATACCCACCGCGCGCCACACCGCGCCAC 3691
QY 718 eValVal----- 720
DB 3692 CGCGGTCTGACGCGCTGTCAACACCAACCACTCACCCACCGACCAACCACTCATGTCGA 3751
QY 721 -----GlyThrIleGlyValThrGlyAlaTyrArgSerAs 733


```

Db      3752  CACCACACCGACCCCGCGCGCTCACCAGGCTCACCAGGCGCCAGACGA 3811
Qy      733  pAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGI 753
Db      3812  ACACCCCGACCGCATCGCGCTCATGAAACCGACACCGCCACACCCCGCTCCCGCTGGC 3871
Qy      753  yGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyr----- 767
Db      3872  CCAACTCGCCACCTCGACACCCCGCTCGGCTCACCACACACCCCTCCACCC 3931
Qy      768  -----IleGlyThrAsnGlyArgGlyIleValTyr-GlyAspIleGlyGlyA 783
Db      3932  CCACCTACCCCGCTCCACACACCCCGCCACCCACACACCCCGCTCAACCCCGACCA 3991
Qy      783  laProSerGlySerPro---SerProSerValSerProSerAlaSerProSerLeuSerP 802
Db      3992  CGCCATCATCATACCGCGGCTCGGCGACCTCGCGGCATCTCGCGCGCACCTGAA 4051
Qy      802  roSerProSerSerSerPro----- 812
Db      4052  CCACCCCGACACCTACCTCTCTCCCGACCCCGCCCGCGCGCGCGCGCGCGCGCA 4111
Qy      813  --SerProSerProSerSerProSerProSerProSerProSerProSerProSerPro 831
Db      4112  CCTCCCTGCGAGCTCGGCGACCCCGCCACCACTCGCCACACCCCTCACCCACATCC 4171
Qy      831  erProSerPro-----SerArgSerProS 839
Db      4172  ACCCTCACCGCATCTTCCACACCGCGCGCACCTCGAGCGGCATCTCTCCAGCGCT 4231
Qy      839  erProSerAlaSerProSerProSerSerSerPro-----S 853
Db      4232  CACCCCGACCGCTCACACCGCTCTCCACCCCAAGCCAAACCGCGCTGGCACCTGCA 4291
Qy      853  erSerSerProSerSerProSerProThrProSerSerPro 868
Db      4292  CCACCTACCCCAAAACCAACCCCTCACCCACTTGTCTCTACTCCA 4338

```

Search completed: October 5, 2005, 21:26:53
 Job time : 627.52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 08:30:02 ; Search time 1843.53 seconds
(without alignments)
3584.301 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENRLTMRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09917376/runat_04102005_164329_26954/app_query.fasta_1.2524
-DB=Published Applications NA -OPM=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -SPART=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09917376 @CGN 1.1 2178 @runat_04102005_164329_26954
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCGAPOP=6 -FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 2: /cgn2_6/ptodata/1/pubna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/ptodata/1/pubna/US05_NEW_PUB.seq:
- 4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:
- 5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq:
- 6: /cgn2_6/ptodata/1/pubna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq:
- 8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq:
- 9: /cgn2_6/ptodata/1/pubna/US09A_PUBCOMB.seq:
- 10: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq:
- 11: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq:
- 12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq:
- 13: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq:
- 14: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq:
- 15: /cgn2_6/ptodata/1/pubna/US10C_PUBCOMB.seq:
- 16: /cgn2_6/ptodata/1/pubna/US10D_PUBCOMB.seq:
- 17: /cgn2_6/ptodata/1/pubna/US10E_PUBCOMB.seq:
- 18: /cgn2_6/ptodata/1/pubna/US10F_PUBCOMB.seq:
- 19: /cgn2_6/ptodata/1/pubna/US10G_PUBCOMB.seq:
- 20: /cgn2_6/ptodata/1/pubna/US10H_PUBCOMB.seq:
- 21: /cgn2_6/ptodata/1/pubna/US10I_PUBCOMB.seq:
- 22: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq:
- 23: /cgn2_6/ptodata/1/pubna/US11A_PUBCOMB.seq:
- 24: /cgn2_6/ptodata/1/pubna/US11_NEW_PUB.seq:
- 25: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq:
- 26: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5134	100.0	2869	11	US-09-917-376-2
2	5134	100.0	2869	15	US-10-155-400-2
3	2487	48.4	9025608	15	US-10-156-761-1
4	2478	48.3	2646	15	US-10-156-761-1845
5	1631.5	31.8	2517	15	US-10-026-994-4
6	1631.5	31.8	2710	15	US-10-026-994-1
7	1570	30.6	9025608	15	US-10-156-761-1
8	1442	28.1	2217	15	US-10-156-761-2561
9	1167.5	22.9	3668	10	US-09-927-827-21
10	1125.5	21.9	2646	18	US-10-395-241-11
11	1124	21.9	2481	18	US-10-395-241-17
12	1117	21.8	2367	18	US-10-395-241-13
13	940	18.3	5698	18	US-10-420-191-1
14	823	16.0	3687	10	US-09-917-384-2
15	823	16.0	3687	10	US-09-917-383-2
16	702	13.7	2289	10	US-09-917-378-2
17	392.5	7.6	1103	20	US-10-653-047-7511
18	365.5	7.1	651	22	US-10-450-763-20760
19	361	7.0	2223	15	US-10-156-761-550
20	346.5	6.7	2706	21	US-10-939-262-11
21	345	6.7	18596	9	US-09-880-107-1590
22	345	6.7	18596	9	US-09-967-768A-119
23	345	6.7	18596	9	US-09-954-531-124
24	345	6.7	18596	9	US-09-954-531-348
25	345	6.7	18596	19	US-10-629-313-140
26	345	6.7	18596	21	US-10-843-641A-1191
27	345	6.7	18596	21	US-10-843-641A-1415
28	345	6.7	18596	21	US-10-843-641A-6264
29	331	6.4	1887	21	US-10-939-262-13
30	330.5	6.4	7407	16	US-10-246-330-3
31	330.5	6.4	7407	17	US-10-282-122A-30151
32	319.5	6.2	9495	19	US-10-669-161-71
33	299.5	5.8	10528	13	US-10-087-192-289
34	297	5.8	32591	17	US-10-085-117-187
35	296.5	5.8	4767	10	US-09-940-316B-28
36	296	5.8	96602	17	US-10-085-117-61
37	293	5.7	4818	10	US-09-940-316B-32
38	292.5	5.7	185548	14	US-10-175-523-62
39	292.5	5.7	185548	24	US-11-099-265-62
40	290	5.6	1232	19	US-10-761-169-17
41	289	5.6	324	17	US-10-282-122A-13296
42	288.5	5.6	4571	10	US-09-940-316B-18
43	288	5.6	288	17	US-10-437-708-111
44	288	5.6	288	17	US-10-395-402-111
45	288	5.6	288	20	US-10-257-199-111

ALIGNMENTS

RESULT 1
US-09-917-376-2
; Sequence 2, Application US/09917376
; Publication No. US2004003834A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOGLYCOLICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1


```
Db 1861 GACCGGGTGAATCCAAAGACTTCTATGCCCTATCAATGGAACCTTCTATCAAGCAGC 1920
Qy 641 AspGlyValThrPheGlnProValalalaGlyLeuProSerSerGlyAlaValGly 660
Db 1921 GACGCGCGGTGACATTCACCGGTCCGCGCGTCTTCCAGCAGCGGTGCCGTCCGT 1980
Qy 661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTlpLeuAlaAAserSerGly 680
Db 1981 GTGATGTTCCACGCGGTGCCGAAAGAGCGGATCTGTGGCTCGCTGCATCGACGCGG 2040
Qy 681 LeuYrHisSerThrAsnGlyLysSerSerTrpSerAlaIleThrGlyValSerSerAla 700
Db 2041 CTTTACCACTCAACCAATGGCGGACAGTGTCTGCAATCACCGCGGTATCTCCGCG 2100
Qy 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720
Db 2101 GTGAACGTGGGATTTGGTAAGTCTCGCGCGCGGTCTATACCCAGCGGTCTTTCTCGTC 2160
Qy 721 GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpVal 740
Db 2161 GGCACGATCGGAGCGGTTACGCGGCGGTACCGCTCCGACGACTGGGACGACCTGGGTA 2220
Qy 741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760
Db 2221 CTGATCAATGATGACACGACCAATACGGAATTCGGGACAGCAATCACCGTGACCCAC 2280
Qy 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyValValTyrGlyAspIle 780
Db 2281 GCGAATTTAGCGCGGTGTACATAGGCACGAACGCGCGTGGAAATGTATACGGGACATT 2340
Qy 781 GlyAlaProSerGlySerProSerProSerProSerValSerProSerAlaSerProSerLeu 800
Db 2341 GGTGTGGCGCGTCCGGATCGCGCTCCGTCGGTGAGTCCGTCGGCTTCGCGCGACCTG 2400
Qy 801 SerProSerProSerProSerSerProSerProSerProSerProSerProSerProSerPro 820
Db 2401 AGCCGAGCCCGAGCCGAGCAGCTCGCCATCGCCGTCCCGTCCCGTCCCGTCCCGTCCCG 2460
Qy 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 840
Db 2461 TCCTCGTCCCGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 2520
Qy 841 SerAlaSerProSerProSerSerProSerProSerProSerProSerProSerProSerPro 860
Db 2521 TCGGCGTCCCGCAGCCGCTTTCGTACCGAGCCCGTCTTCGTACCGCTTCCTCGTCCGCG 2580
Qy 861 SerProThrProSerSerProValSerGlyValLysValGlnTyrLysAsnAsn 880
Db 2581 AGCCCAACGCCGCTCGTCCGCGGTGTCGGGTGGGGTGAAGGTGCAGTATAGAAATAT 2640
Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
Db 2641 GATTCCGCGCGCGGTGATATACAGATCAAGCCGGGTTTCAGGTGGTGAATACCGGGTCG 2700
Qy 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyLysSer 920
Db 2701 TCGTGGTGGATTTGTGACGCGGTGACGCGTCCGCTACTGGTTTACCCCGGATGTGGCTCG 2760
Qy 921 SerThrLeuValTyrAsnCysAspTrpAlaIleGlyCysGlyAsnIleArgAlaSer 940
Db 2761 TCACACTGGTGTACAACTGTGACTGGGCGGGGATCGGGTGGGAATATCCGCGCTCG 2820
Qy 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 2821 TTCGCTCGGTGAAACCCGCGAGCGCGACGCGGACACTTACTTCGAC 2868
```

RESULT 2

US-10-155-400-2

; Sequence 2, Application US/10155400

; Publication No. US2003010898A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

```
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2
```

Alignment Scores:

Pred. No.:	0	Length:	2869
Score:	5134.00	Matches:	956
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.98%	Indels:	0
DB:	15	Gaps:	0

US-09-917-376-1 (1-957) x US-10-155-400-2 (1-2869)

Qy	1	MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgLeuValSerLeu	20
Db	1	ATGGATCGTTCGAGAACATCCGTCTGACTATGAGATCAGACGATTGGTATC	60
Qy	21	AlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIle	40
Db	61	CGCGCCACTCGTTCGTCGCGCGCTCTGGAGTTCTGCCCATCGCGATAACG	120
Qy	41	AlaSerProAlaHisAlaIleThrGlnProTyrThrTrpSerAsnValAlaIle	60
Db	121	GCTTCTCTCCGCGACGCGCGGACGACTCAGCCGTACACCTGGAGCAACG	180
Qy	61	GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu	80
Db	181	GCGCGCGCTTGTGACCGGATCGTCTCAATGAAGTGCACCGGAATTCGTAC	240
Qy	81	ArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle	100
Db	241	CGGACGACATCGGGGGATGTATCGATGGATGCCCAACGCGCGGTGGATCC	300
Qy	101	LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla	120
Db	301	CTGGATTGGGTGGATGGAAACAATTCGGGGTCAACGCGGTCGTACAGATT	360
Qy	121	ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSer	140
Db	361	CCGATCAATACTAACAGGTATGGCGCGCTCGAATGTACACCAACAGCTGG	420
Qy	141	AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIle	160
Db	421	AACGACGGAGCGATTCTCCGTCGTATCAGCGCGCAACGTTGGCAATAAC	480
Qy	161	ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla	180
Db	481	CCGTTCAAGCTTGGCGGCAACATSCCGCGCGTGGGAATGGCGGAGCGGCT	540
Qy	181	ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp	200
Db	541	CCAAACAATGACAACTTCTGTATTTTCGGCGCCCGCGAGCGCGCAAGGCT	600
Qy	201	ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGly	220
Db	601	ACAGATTCGCGCGGACCTGTTCCAGATGACGAACCTTTCCGACGCTAGGCA	660

Qy	221	AlaasnProthrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla	240	Db	1741	GACGGCTCTCGTTTCGTCTGGGCTCCCGCGCATCCCGGTACGCTTGTTGTGTACGAGTC	1800
Db	661	GCAATCCCACTGACACGACCGGCTATCAGAGCGATATTCAAGGCGTGGTGGTGGCT	720	Qy	601	GlyPheGlyAsnSerTrpAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer	620
Qy	241	PheAspIysSerSerSerLeuGlyGlnAlaSerIysThrIlePheValGlyValAla	260	Db	1801	GGATTTGGCAACTCTCTGGGCTGCTTCGCAAGGTGTTCGCGCAATGCCAGATCCGCTCA	1860
Db	721	TTGACAAAGTCTTCGTTCATCGCTCGGGCAAGCGAGTAAGACCATTTTGTGGCGTGCGG	780	Qy	621	AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr	640
Qy	261	AspProAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro	280	Db	1861	GACCGGTGAATCCAAAGACTTCTATGCCCTATCCAATGGAACCTTCTATCGAAGCAGC	1920
Db	781	GATCCCAATAATCCGGTCTTCTGAGCAGACACGGCGCGGACGTGGCAGCGGTGCGG	840	Qy	641	AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly	660
Qy	281	GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu	300	Db	1921	GACGGCGGTGACATCCAAACCGGTCGCGCGGTCTTCCGAGCAGCGGTGCGGTGCT	1980
Db	841	GGTGGCGGACCGGCTTCATCCCGCACAAAGGGCGTCTTTGACCCGGTCAACACAGTGTCTC	900	Qy	661	ValMetPheHisAlaValProGlyLysGlyAspLeuLeuTrpLeuAlaAlaSerSerGly	680
Qy	301	TyrIleAlaThrSerAsnThrGlyProTyrAspGlySerSerGlyAspValTrpIys	320	Db	1981	GTCAATGTTCCACGGCGGTGCTGGAAAAGAGGCGATCTGTGGCTCGCTCATCGAGCGGG	2040
Db	901	TATATTGCCACCAAGATAACGGTGGTCCGTATGACGGAGCTCCGGCGAGCTCTGGAAA	960	Qy	681	LeuTyrHisSerThrAsnGlyLysSerTrpSerAlaIleThrGlyValSerSerAla	700
Qy	321	PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla	340	Db	2041	CTTTACCACTCAACCAATGGCGGACGAGTTGGTCTGCAATCACCGCGGTATCTCTCGCG	2100
Db	961	TTCTCGGTGACCTCCGGACATGACCGGAATCAGCCGCGTACCTTCGACGGACACGGCC	1020	Qy	701	ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal	720
Qy	341	AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet	360	Db	2101	GTGAACGTGGGATTTGGTAAGTCTGCGCCCGGTGCTCATACCCAGCCGTCTTTGTGCTC	2160
Db	1021	AACGACTACTTTGGTTACGGGGCTCACTATCGACCGCCAGCACCCGACCATTAATG	1080	Qy	721	GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpVal	740
Qy	361	ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly	380	Db	2161	GGCACGATCGAGGGGCTTACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTA	2220
Db	1081	GTGGCAACCCAGATATCGTGGTGGCGGACACCATTAATCTTTTCGGACCGACCGCGGT	1140	Qy	741	LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis	760
Qy	381	AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal	400	Db	2221	CTGATCAATGATGACACACCAATACGGAATTTGGGACACAGCANTCACCGGTGACCAC	2280
Db	1141	GCAGCGTGGACCGGATCTGGGATTTGACGAGTATATCCCAATCGAAGCTTCGGATATGTG	1200	Qy	761	AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle	780
Qy	401	LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProValPro	420	Db	2281	CGCAATTTACGGCGGTGTACATAGGCACGAACGCGCGTGAATTTATACGGGGACATT	2340
Db	1201	CTTGACATTTTCGGCGGAGCCTTGGCTGACCTTCGGCGTACAGCCGAATCTCCCGTACCC	1260	Qy	781	GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu	800
Qy	421	SerProLysLeuGlyTyrTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg	440	Db	2341	GGTGTGCGCGCTCCGGATCGCGCTCTCGGTGAGTCCGTCCGCTTCGCGAGCCCTG	2400
Db	1261	AGTCCGAAGCTCGCTCGATGATGAAGCGATGCAATCGATCGCTTCAACTCTGATCGG	1320	Qy	801	SerProSerProSerProSerSerProSerProSerProSerProSerSerPro	820
Qy	441	MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer	460	Db	2401	AGCCCGACCCGAGCCGAGCAGCTCGCCATCGCCGTCCCGCTCGCGAGCTCGAGTCCA	2460
Db	1321	ATGCTCTACGGAACAGCGCGACGTGTGTACGCAACAAATGATCTCACGAATGGGACTCC	1380	Qy	821	SerSerProSerProSerProSerProSerProSerProSerProSerProSerPro	840
Qy	461	GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp	480	Db	2461	TCCTCGTCGCGCTCTCGTCGCGCTCACCATCGCGAGTCCGTCTCGGTCCGTCAACA	2520
Db	1381	GGCGGCGAGATTCATATCGCGCCGATGTTCAAGGATTTGGAGGAGACGGCGGTAAACGAT	1440	Qy	841	SerAlaSerProSerProSerSerProSerProSerProSerProSerProSerPro	860
Qy	481	LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe	500	Db	2521	TCGGGCTCGCGAGCCGCTCTTCGTACCGAGCCCGCTCTTCGTACCCGCTTCCTCGTCGCG	2580
Db	1441	CTCATACGCGCGCTCTGGCGCCCGCTCATACGCGCTCTCGGAGACTTCGGCGGCTTC	1500	Qy	861	SerProThrProSerSerProValSerGlyValLysValGlnTyrLysAsnAsn	880
Qy	501	ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr	520	Db	2581	AGCCCAACGCGCTCTGTCGCGGTGCGGTGGGTGAAGTGCAGTATTAAGATAAT	2640
Db	1501	ACCCACGCGGAGTTACTGCGCGTCCATCGACGATCTTCAGCTCACCGGTGTTCAACGACC	1560	Qy	881	AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer	900
Qy	521	GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe	540	Db	2641	GATTCGGCGCGGTGATAATCAGATCAAGCCGGGTTTCGAGGTGTGTAATCCGGGTGCG	2700
Db	1561	GGCACCGCGTCAATGCGGAATGAATCCGTCGATCATCTGTCGCGCTGGAGATTTC	1620	Qy	901	SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer	920
Qy	541	AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn	560	Db	2701	TCGTCCGTGGATTTGTCGACGCTGACGCTGCGGTGCTGCTGCTTACCCCGGATGCTGCTG	2760
Db	1621	GATCATCGACCAACCAACGACAGCGACGCTCGGTCTCGACAGACGGCGGCAAGAAC	1680	Qy	921	SerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer	940
Qy	561	TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla	580	Db	2761	TCGACACTGGTGTACAACCTGTACCTGGCGGCGATCGGGTGTGGGAATATCCCGGCTCG	2820
Db	1681	TGGTTCCAAAGCAGCAACCTGGCGGGTGGACGACGGCGGCGACCGTCCGCGCATCGGCC	1740	Qy	941	PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln	956
Qy	581	AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal	600				


```
Db      2821 TTCGGCTCGGTGAACCCGGGACGCCGACGGCGGACACCTACCTGCAG 2868
RESULT 3
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HAITORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.:          2,57e-161      Length:          9025608
Score:             2487.00         Matches:          490
Percent Similarity: 64.77%         Conservative:     135
Best Local Similarity: 50.78%       Mismatches:      248
Query Match:       48.43%           Indels:          93
DB:                15              Gaps:            18

US-09-917-376-1 (1-957) x US-10-156-761-1 (1-9025608)
QY      9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28
Db      2277388 CTCACGGTCGGAAGAACCGCATCTTC-----ACGGCCGTGCTCGGCGTG 2277432
QY      29 AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46
Db      2277433 GCGGCGGTCGTCGCCCGCGGCACCCCGCGGCACCTGCGCGGCACCTGCGCGCGGCGGCGAGC 2277492
QY      47 AlaThrThrGlnProTyrThrTrpSerAenValAlaIleGlyGlyGlyPheValAsp 66
Db      2277493 ATGCGCGCGGACAGTACAGCTGGAAGAACCGCGCGTGCAGCGCGCGCGCTTCGTCCCC 2277552
QY      67 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86
Db      2277553 GGCATGCTCTTCAACCGCTCCGAGAGAACCTCGCTACGCGGCACCGACATCGCGCGC 2277612
QY      87 MetTyrArgTyrAspAlaAlaAlaGlyA-gTTrpIleProIleLeuAspTrpValGlyTyr 106
Db      2277613 GCCTACCGCTGGCGCGAGTCCTCGAAGACCTGAGACCGCGCTGCTCGACTCGGTGCGGTGG 2277672
QY      107 AsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126
Db      2277673 AGCGACTGGGGGACACCGGGTGTGAGCCTCGCTCCGACTCCGTCGACCCGGAACAG 2277732
QY      127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTyrAspProAsnAspGlyAlaIleLeu 146
Db      2277733 GTGTACCGGGCGCTCGGCAGTACACAGACAGTGGGACCGGGCAACGGTGGCGGCTC 2277792
QY      147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166
Db      2277793 AGTCCCGCGACACCGGGCGCGAGCTGGCAGAAAGACCGACCTGCCCTTCAAGCTGGGCGGG 2277852
QY      167 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 186
Db      2277853 AACATGCGCGGCGCGGGCATGGCGAGCGGCTCGCGGTGCACCCGACAGAACAGAGCGGTG 2277912
QY      187 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 206
Db      2277913 CTGTATCTCGGCGCGCCACGCGGCAAGGGGCTGTGCGGTTCGAGGACTCGGGGCGCTCC 2277972
QY      207 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAlaProThrAspThr 226
Db      2277973 TGGTCGAGGTCAACGACTTCCGAAACGTCGCGCACCTCTGTGAGGACGCGACCGACACG 2278032
QY      227 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 246
Db      2278033 AGCGGGTACGCGTCCGCAACACGAGGATCGTGTGGTCACTTCGACGAGTTCGACGGG 2278092
QY      247 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 266
Db      2278093 TCGCGGGGAGTCCACGCGGAGCGGTGAGTCTGGGTTCGCGGACAGGACAACTCCGTC 2278152
QY      267 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286
Db      2278153 TATCGCTCCACGACGCGGCGGACCTGGTCCGCGTGGCGGACCGACCGACCGGCGCAT 2278212
QY      287 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306
Db      2278213 CTCGCCACAAAGGGCGTGTGGACGCGGGAACGGGTGTGTACCTTCGGGTACAGGCAC 2278272
QY      307 ThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysPheSerValThrSerGly 326
Db      2278273 AAGGGCGGACCGTACGACGCGCGCAAGGGACACTGTGGCGGTGACACGACGAGACCGGG 2278332
QY      327 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAlaAspTyrPheGlyTyr 346
Db      2278333 ACCTGGACGAACATACGCCGGTCCGCGGAGCGCGACACC-----TACTACGGCTTC 2278383
QY      347 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366
Db      2278384 AGCGGCTGACCTGGACCGGACATCCGGGACCGGTGATGGCGACTGGGTACACGACGAGAC 2278443
QY      367 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 386
Db      2278444 TGGTGGCGGACACGACGACTCTTCGCTCCACGACAGCGCGGCGGACCTGGACGAGGCC 2278503
QY      387 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 406
Db      2278504 TGGGACTACACCTCGTATCCGAGCGCTCGAACCGCTTCACCATGGATGTCTCGTCTCG 2278563
QY      407 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTyr 426
Db      2278564 CCCTGGCTACCTGGGAGCGAACCCCGCACCGCGGACGAGACCCCGAAACTCGCGCTGG 2278623
QY      427 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 446
Db      2278624 ATGACCGAGTCCCTGGAGATCGACCCGTTGCATCCGCGCGCATGATGATGACGAGACGG 2278683
QY      447 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 466
Db      2278684 GCGACGGTCTACGGCAGCGCAACCTGACGAACTGGGACGCGGAGAGCCAGTTCACCATC 2278743
QY      467 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 486
Db      2278744 AAGCCGATGGCGGGCGCTGGAGAGACGCGCGCTCAACGACCTCGCTCGCTCCCTCC 2278803
QY      487 ---GlyAlaProIleLeuSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal 505
Db      2278804 GCGCGCGCCGAGTGTTCAGCGCGCTCGGTGACATCGGCGCTTCCCGGCGCTCCCGGACCTC 2278863
QY      506 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 525
Db      2278864 ACCACGGTCCGCTGCTGATGTATACATCGTCCGCGAACTTCACCAAGACGACCGACCTCG 2278923
```


QY 526 TyrAlaGluLeuAenProSerIleIleValArgAlaGlySerPheAaspProSerSerGln 545
 Db 2278924 TACCGCGAGACCGCCCGGACCGGTGTGGGTGCGCAATCTCGAC-----TCGGGT 2278977
 QY 546 ProAenAaspArgHisValAlaPheSerThrAaspGlyGlyLysAenTrpPheGlnGlySer 565
 Db 2278978 CCG-----CATGTGCGGTCTTCGACGCGACAAACGCGCGCAACTGTGTTCGCGGCGCG 2279028
 QY 566 GluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAaspGlySerArgPhe 585
 Db 2279029 GACCTTCGCGGGTCAGCGGGGTGGGACGTCGCGCGGTTCGACGCGGAGTGCCTTC 2279088
 QY 586 ValTrpAlaProGlyAaspProGlyGlnProValValTyrAlaValGlyPheGlyAasnSer 605
 Db 2279089 GTGTGGAGCCCGCGGCGACCGG-----GTGCGAGTACACACCGGGTTCGGACCTCG 2279142
 QY 606 TrpAlaAlaSerGlnGlyValProAlaAasnAlaGlnIleArgSerAaspValAasnPro 625
 Db 2279143 TGGTCGGGTTCGCGGGCTCCCGCGCGGGGATCGTTCGAGTCCGACCGGTTCGACCGG 2279202
 QY 626 LysThrPheTyrAlaLeuSerAasnGlyThrPheTyrArgSerThrAaspGlyGlyValThr 645
 Db 2279203 AAGACCTTCTACGGCTTCAAGTCCGCGACAGTTCTACGTCACTTCGACGCGCGGCGGAC 2279262
 QY 646 PheGlnProValAlaAlaAa---GlyLeuProSerSerGlyAlaValGlyValMetPheHis 664
 Db 2279263 TTCACGGGTCCGCGGGCGCGGCTGCGGAGCGCGACAGC-----GTGCGCTTCAAG 2279316
 QY 665 AlaValProGlyLysGluGlyAaspLeuTrpLeuAlaAaSerSer----- 679
 Db 2279317 CGGTGCGCGGACGAGCGGCGACATCTGGTGGCGGCGCGGACGCGCGGTAC 2279376
 QY 680 GlyLeuTyrHisSerThrAasnGlySerSerTrpSerAlaIleThrGlyValSerSer 699
 Db 2279377 GGGCTGTGGCACTCGACGCGCGCGCGCGCTTCCACCAAGCTCGCCACCGTTCGACCG 2279436
 QY 700 AlaValAenValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719
 Db 2279437 GCGGACACCATCGGCTTCGCGAAGCGCGGACCGCGCTCGTACCAACGCTCTACACC 2279496
 QY 720 ValGlyThrIleGlyValThrGlyAlaTyrArgSerAaspCysGlyThrThrTrp 739
 Db 2279497 AGCGGAAGATCGCGGTGTGGCGGCGATCTTCGGTTCACCGACAGGGCGCGAGCTGG 2279556
 QY 740 ValLeuIleAenAaspAaspGlnHisGlnTyrGlyAasnTrpGlyGlnAlaIleThrGlyAasp 759
 Db 2279557 ACCCGCGTCAACGACGATGCCACCACTGGGGTTCGACGCGGCGCGCGATCACCGGTGAC 2279616
 QY 760 HisAlaAenLeuArgArgValTyrIleGlyThrAenGlyArgGlyIleValTyrGlyAasp 779
 Db 2279617 CCCAGGGTCTACGGGCGGTGTATGTGTGTCGACGAAACGCGGCGGGATCGTCTACCGCGAC 2279676
 QY 780 IleGlyGlyAlaProSerGly-----SerProSerProSerValSerProSerAla 796
 Db 2279677 ACCCGGGCTCTTCGAGCGCGCGGTACGAGAGCCGCGACCGCGCGCGCGCTGACGGTG 2279736
 QY 797 SerProSerLeuSer-ProSerProSerProSerSerSerProSerProSerProSer 815
 Db 2279737 ACATACAGGATCACCAACAGTGGTGGCGCGCTTCCAGGCGGATGTACAGCTCGCCCAAC 2279796
 QY 815 rProSerSerSerProSerSerSer----- 824
 Db 2279797 ACCCGGTTCGACCGCGCTGGGACCGGCTCGTGGGCTGCTGCTGCGTACCGGCGGAG 2279856
 QY 824 oSerProSer----- 827
 Db 2279857 GTACCCAGCTGTGAACGCTCTGACGCGGAGCGCGGTTCGCGGGTGTACCGCAGGAAC 2279916
 QY 828 -----ProSerPro----- 830
 Db 2279917 CTGGCTGGAACGGGAGGTGGCGCGGCTCTCGGTGAGCTTCGGGTTCACGGGAGC 2279976

QY 831 -----SerProSerProSerArgSerProSerProSerAla-----Se 843
 Db 2279977 TGGTCGGGATCCATATGGCAGACCGACCGCTTCAAACTCGGCGGTGAGAAGTGCACCGGTC 2280036
 QY 843 rPro-----SerProSerSerSerProSerProSerProSerProSerProSe 861
 Db 2280037 GCCTGATCGAAGCGCGCA-CGCTCGCTCATCCGAGGCGCGGCGCGGCGCGCG 2280095
 QY 861 rProThrProSerSerSerProValSerGlyGlyVallyValGlnTyrLysAasnAas 881
 Db 2280096 CCTCCGGCTCTCCACGCGGTGTAGACCGTGGCTG-----GGCAA 2280137
 QY 881 pSer-AlaProGlyAaspAasnGlnIleLysProGlyLeuGlnValAasnThrGlySerS 901
 Db 2280138 CTCCCGCTCTGTGTGAGGCTTCCGCGCAGTCCGGCTCGGCTCGGCGCGGCGAGTA 2280197
 QY 901 erSerValasp 904
 Db 2280198 CCGCGCACGAC 2280208
 RESULT 4
 US-10-156-761-1845
 ; Sequence 1845, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIORITY APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1845
 ; LENGTH: 2646
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2646)
 US-10-156-761-1845
 Alignment Scores:
 Pred. No.: 2,25e-164 Length: 2646
 Score: 2478.00 Matches: 473
 Percent Similarity: 64.07% Conservative: 128
 Best Local Similarity: 50.43% Mismatches: 227
 Query Match: 48.26% Indels: 110
 DB: 15 Gaps: 14
 US-09-917-376-1 (1-957) x US-10-156-761-1845 (1-2646)
 QY 23 ThrAlaSerPheAlaValAlaAlaLeu-----GlyValLeuProIleAlaIleThr 40
 Db 22 ACGGCGGTCTCGCGTGGCGCGGTCTGCCCGCGGACCCCGCGGACCTGGCGCGG 81
 QY 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAasnValAlaIleGly 60
 Db 82 AGCGCCCCACCGCGACGATCCCGCGGACAGTACAGCTGGAAGAACCCCGCGTCCG 141
 QY 61 GlyGlyGlyPheValAaspGlyIleValPheAasnGluGlyAlaProGlyIleLeuTyrVal 80
 Db 142 GGCGCGCGGTCTGTCGCCGCGCATCGTCTTCAACCGCTCCGAGAGAACCTCGCTACGCC 201
 QY 81 ArgThrAaspIleGlyGlyMetTyrArgTrpAaspAlaAasnGlyArgTrpIleProLeu 100

202	Db	 CGCACCGA	CATCGCGCGCGCTACCGCTGGCGCGAGTCTCTCGAAGACCTTGGACCGCGCTG	261
101	Qy	LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp	120	
262	Db	CTGACCTCGGCTGGCTGGAGCGACTCGGGGCGACCGGGTGTCGTGAGCCTCGCCTCGCAC	321	
121	Qy	ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro	140	
322	Db	TCGCTCGACCCGAACAAGGTGTACCGCGCGCTCGGCACGTACACGAACAGCTGGGACCCG	381	
141	Qy	AsnAspGlyAlaIleLeuArgSerAspGlnGlyAlaThrTrpGlnIleThrProLeu	160	
382	Db	GGCAACGGTCCGCTGCTCAGTCCGGCGACCGGGGGCGAGCTGGCAGAAAGACCGACCTG	441	
161	Qy	ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp	180	
442	Db	CCCTTCAAGCTGGCGGGGGAACATGCCGGGCGGGGTCATGGCGGACCGCTCGCGGTCGAC	501	
181	Qy	ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer	200	
502	Db	CCGAAcAGGAACAGCGTGTGTATCTCGGCGCGCCACCGGCAAGGGCTGTGGCGGTGCG	561	
201	Qy	ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle	220	
562	Db	ACGACCTCGGGGCGCTCTCGGTGCGAGGTCAcCGACTTCCCGAAACGTTCGCGACCTACGTG	621	
221	Qy	AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla	240	
622	Db	CAGGACCGCACCGACACGACGGGTACGCTCGACAAcCAGGCGCATCTGTGTGGGTCAcC	681	
241	Qy	PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla	260	
682	Db	TTTGACAGTTCGACGGGGTCCCGGGGAGCTCCACGGGAGCGGTGTACGTTCGGGTGCC	741	
261	Qy	AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro	280	
742	Db	GACAAcGACAACTCCGTCTATCGCTCCACGACCGGGCGCGACCTGCTCCCGCTGCC	801	
281	Qy	GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu	300	
802	Db	GGCCACGCCACCGGCATCTCGGCCACAAcAGGGGGTGTCTGACCGCGCGAAGCGGTGTCTG	861	
301	Qy	TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys	320	
862	Db	TACTCTCGGTACAGCAGCAGGGCGGACCGTACGACCGCGCGCAAGGACAGCTGTGGCGG	921	
321	Qy	PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla	340	
922	Db	TACACGACGAAGACCGGACCTGGACGAACATCAGCCCGGTCTCGCGAGGCGCACACCC--	978	
341	Qy	AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet	360	
979	Db	-----TACTACGGCTTCACGGCGGCTACCGTGGACCGGACGATCCGGGACCGGTGATG	1032	
361	Qy	ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly	380	
1033	Db	CGGACTGCGTACAGCTCTGGTGGCCCGGACACGCGAGCTCTTCGCTCCACGGACAGCGGC	1092	
381	Qy	AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal	400	
1093	Db	GGCACCTGGACGAAGGCTGGGACTACACCTCGTATCCGAGCGCTCGAACCCGCTTCACC	1152	
401	Qy	LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro	420	
1153	Db	ATGGATGTCTCGTCTCCCGCTCGCTCACTGGGAGCGGAACCCCGCACCGCCCGACGAG	1212	
421	Qy	SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg	440	
1213	Db	ACCCCGAACTCGGCTGGATGACCGAGTCCCTGGAGATCGACCGCTTCGACTCCCGCGGC	1272	
441	Qy	MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer	460	


```
Db 1219 GACCCACCGACGACCACTGGCTCTACGCGACCGGAATGACAATCTTTGGCGCCAC 1278
Qy 454 AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValIleGlyLeu 473
Db 1279 GATCTCAGCAACTGGGACACGCGCACCAATGTGTCAATCCAATCATCGCAGACGCGCATC 1338
Qy 474 GluGluThrAlaValAsnAspLeuIleSerProSerGlyValAlaProLeuIleSerAla 493
Db 1339 GAGGAATTTCTCGCTCAGGACCTGGCTTGCACCGCGGAGCGAGCTATTTGGCCGCA 1398
Qy 494 LeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSer 510
Db 1399 GTGAGAGACGACCAACGGCTTACCTTTGCCAGCAGAAACGACCTCGGACATCGCGCAG 1458
Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
Db 1459 ACGGTCTGGGCAACGCCACATGGGCCACCTCGACGAGCGTGCATPACGCGCGGAACCTCG 1518
Qy 531 ProSerIleIleValArgAlaGlySerPheAspProSerPheAspProSerGlnProAsnAspArgHis 550
Db 1519 GTCAAGAGCGTGTCCGCTCGCGTCCGCAACACCGCGCGCACCAACAG-----1563
Qy 551 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal 570
Db 1564 GTGGCCATCTCTCGACGCGCGCGGCGGACGTGAGCATCGACTACGCGCGCGCACGCTCC 1623
Qy 571 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 590
Db 1624 ATCAACGGCGGACGCGTGGCTTATTCGGCGCGGCGGACGACGATCTCTGTGTCGACCGCC 1683
Qy 591 AspProGlyGlnProValValTyrAlaValAlaGlyPheGlyAsnSerTrpAlaAlaSerGln 610
Db 1684 TCGTCCGCG-----GTGACGCGCTCGCAGTTTCCAGGCGCAGCTTTCCTCGCTCG 1734
Qy 611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 630
Db 1735 AGCTGCGCGCGCGCGCGCTATCGCTCGGACGAGACGATCGCGGATATCGTCTCACCGCC 1794
Qy 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 650
Db 1795 GGTCTCGGATCGACCTTTTACGTACGACGAGGACACCGCGCAGCAGCTTC-----ACG 1845
Qy 651 AlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668
Db 1846 CGCGGG---CCCAAGCTGGCAGCGCAGGACGATCGCGGATATCGTCTCACCGGACC 1902
Qy 669 LysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688
Db 1903 ACGCGGGCAGCTGTGTATGTCGACCGACGTCGCGCATATTCGCTCCACAGACTCGGGC 1962
Qy 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707
Db 1963 ACAGCTTTGGCAAGTCTCCACCGCGCTGACCAACACCTACGATCGCGCTGGGTG 2022
Qy 708 SerAlaProGlySerSerTrpProAlaValPheValAlaGlyThrIleGlyValThr 727
Db 2023 GGCTCA---GGCTCGAAGTGG---AACCTGTATGCTTCGGCAC-----GGCCGCTCA 2070
Qy 728 GlyAla-----TyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspApp 745
Db 2071 GGGGCTCGGCTCTACGCCAGTGGAGACAGCGCGCGCTCTCGGACGACATCCAGGGCTCC 2130
Qy 746 GlnHisGluTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765
Db 2131 CAGGGTTCGGCTCCATCGACACCAAGGTTCGCGGACGCGGACGCGCGCGGCGCA 2190
Qy 766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyr-----GlyAspIleGlyGlyAla 783
Db 2191 GTCTAGTGGGACCAACACGCGCGGGCGTCTTTTACGCTCAGGGAACCGTCGCGCGCGC 2250
Qy 784 ProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSer 803
```

```
Db 2251 ACGGGCGGAGCT-----TCCTCGTCGACCAACGACGAGC 2283
Qy 804 ProSerProSerSerSerProSerProSerProSerProSerProSerProSerSer 823
Db 2284 AGCAGAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGGTTGTATCCACG 2343
Qy 824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSerAlaSe 843
Db 2344 ACCCGGGCTTCGACGGTGACTTCGTGAGGACGACGCTCGGCGCGCTCCACGCGGTCA 2403
Qy 843 rProSerPro 846
Db 2404 GGGGTCGCCG 2413
RESULT 6
US-10-026-994-1
; Sequence 1, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Yao, Michael
; TITLE OF INVENTION: EGYI Endoglucanase and Nucleic Acids
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-994-1
Alignment Scores:
Pred. No.: 6,15e-105 Length: 2710
Score: 1631.50 Matches: 353
Percent Similarity: 57.23% Conservative: 130
Best Local Similarity: 41.82% Mismatches: 297
Query Match: 31.77% Indels: 64
DB: 23 Gaps: 23
US-09-917-376-1 (1-957) x US-10-026-994-1 (1-2710)
Qy 25 SerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
Db 50 TCTCGAGTCTTGGCTTGTCTTGGGGCGGCTC-----ATCCCTGCC 91
Qy 45 HisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
Db 92 CATGCTGCC-----TTTTCATGGAAGAACGTCAGCTCGGCGCGCGCGCGC 139
Qy 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 140 TTGTCCTCCCGCATCATCTTCATCCCAAGACAAAGAGCGTAGCATATGCACCAACAGAT 199
Qy 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 200 ATTGGCGGGCTGTACCGCTCAAC---GCCGACGACTCATGACCGCGCTCAGGATGG 256
Qy 104 Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 257 ATTGCTGATAATGCGGCTGCGCAACTGG-----GGCATCGACGCTTGTGGCTT 307
Qy 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
Db 308 GATCCGAGGACGATCAAAAGGTGTATGCCGAGTCGCGCATGTATACGAACAGCTGGAT 367
Qy 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 368 CCGAGTAATGAGGCCATCATTCGCTCGTGCAGACCGCGCGGCGGCTGCTTCACCAAC 427
```


US-10-156-761-1/c
 ; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Alignment Scores:
 Pred. No.: 6,1e-97 Length: 9025608
 Score: 1570.00 Matches: 393
 Percent Similarity: 49.25% Conservative: 131
 Best Local Similarity: 36.94% Mismatches: 346
 Query Match: 30.57% Indels: 194
 DB: 15 Gaps: 38

US-09-917-376-1 (1-957) x US-10-156-761-1 (1-9025608)

QY 4 SerGluAsnIleArgLeuThrMetArg-----SerArgArgLeuValSerLeu 19
 DB 3162026 TCTCCAGACGAGGTATGTCATGCGACGCCGCCGCCGAGCAGACCGTCC----- 3161973
 QY 20 LeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIle 39
 DB 3161972 CTCGCGGGACCGCGCG-----GCCGCGCGCTCACGGCGTCC-----GCC 3161928
 QY 40 ThrAlaSerProAlaHisAlaAlaThrThr-----GlnProTyrThrTrpSerAsn 56
 DB 3161927 GTGGGGTCTCCAGCGCAGCCGCCGAAACACCGCGCGCGCTCGTACGCTGGCGCAAC 3161868
 QY 57 ValAlaIleGlyGlyGlyPheValAlaAspGlyIleValPheAsnGluGlyAlaProGly 76
 DB 3161867 GCCGTATCGGGGACCGCGCTTCGTACCGCGGTCTCTCCACCCCTCCGTACCGGT 3161808
 QY 77 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 96
 DB 3161807 CTCGCTACCGCGGACCGACATCGCGCGCGCTACCGCTGGGACGCGCGCGCGCC 3161748
 QY 97 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSer 116
 DB 3161747 TGGACCCCGCTCATCGACACCTCGGCTGGGACGACTGGAACCTCTCGCGCTCGAGGCG 3161688
 QY 117 IleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsn 136
 DB 3161687 ATGGCGGTCCAGCCACCGACCGCGCTCTACCTCGCGCTGGGACCTACGCGCCAG 3161628
 QY 137 SerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln 156
 DB 3161627 TCGTGGCGGGACAC-----GGCGGGTCTCGCTCGAGGACCGCGCGCCACCTGGACC 3161571
 QY 157 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176

DB 3161570 CGCACCGACCTGACCGTGAAGCTCGCGCGCACAGAGCGCGCGCGCGGTGAGCGA 3161511
 QY 177 LeuAlaValAspProAsnAsnAspIleLeuTyrPheGlyAlaProSerGlyIleGly 196
 DB 3161510 CTCCTCGTCGACCGCGCGCAGCAGCAGCACCTCTGGCTGGGACG---CGGACGACGGG 3161454
 QY 197 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 216
 DB 3161453 CTGCTCAAGTCGACCGCGCGCGCGCTTGGGACGCGCGACCGCTTCCCG----- 3161400
 QY 217 GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal 236
 DB 3161399 -----GCCGAGCGCGAACCTCTCTCGCGCGCGGAGTCC 3161370
 QY 237 ValTrpValAlaPheAspLysSerSerSerSerSerSerSerSerSerSerSerSer 256
 DB 3161369 GTGTTT-----CTCGTCCCGCGCGCGCGCGCGCTAC 3161337
 QY 257 ValGlyValAlaAspProAsnAsnPro-----ValPheTrpSerArgAsp 271
 DB 3161336 GCCGCTGGGTGACCGCGCAGCGCACCTCGGGCAGCGCGACCTGTACCGCAGCGCGC 3161277
 QY 272 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe-----IlePro 288
 DB 3161276 ---GGCAGCAGCTGGGGCGCGTCCCGCGCGCGCTCGCGCACCTCGCGCAAGTCCCG 3161220
 QY 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308
 DB 3161219 CTCGCGCGCGGTACGACACGACCGCGCGCTGTACGTACGTACGCGCAGCGCACCC 3161160
 QY 309 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrTrp 328
 DB 3161159 GGCGCGGGCGCGCAGTCCGACCGCAGCGTGCACAGCTGGTACCGCCACCGGGACGTGG 3161100
 QY 329 ThrArgIleSerProVal-----ProSerThrAspThrAlaAsnAspTyrPhe 344
 DB 3161099 ACGAGGTACCCCGGTGAAGCGCGCGCGCAGCAGCAGCGCGTCCGGGACACCTTC 3161040
 QY 345 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 364
 DB 3161039 GCCTACGGCGGGTCCGCTCGACCGCGCGCGCGCGCGCGCTCGCTCCACCAAC 3160980
 QY 365 IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThr 384
 DB 3160979 AACCGCTGGCGCGCAGCGCGACCGGTCTCCGGTCCACGCGCGCGCGCGCTGACG 3160920
 QY 385 ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 404
 DB 3160919 TCCCTCAAGAC-----GCCGCGCGTTCGACGTGTCC 3160887
 QY 405 AlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeu 424
 DB 3160886 GAGACTCCCTTCCTCGACTGGGCGCGCGAC-----AAGCGGAAGTTC 3160845
 QY 425 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 444
 DB 3160844 GCGTGTGTGATCCAGCGCTCGCGGTCCGCGTCCAGCTCCAGCAGCGCTGTGTACGGG 3160785
 QY 445 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIle 464
 DB 3160784 ACCGGCGGACCTCTACCGCACCGCGACCTCAAGGCTGG----- 3160743
 QY 465 HisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro 484
 DB 3160742 -----GCACCGCGGATCCCGCGCTCGAGGAGAGCGCGCTCGGCCAACCTGATCGGCC 3160689
 QY 485 ProSerGly---AlaProLeuIleSerAlaIleuGlyAspLeuGlyGlyPheThrHisAla 503
 DB 3160688 CCGGTCCGGGAGGACACCTGATCAGCGGACTCGGGGACATCGGTGTGATGTACCGAG 3160629
 QY 504 AspValThrAlaValProSerThr---IlePheThrSerProValPheThrThrGlyThr 522


```
Db 3160628 CGGCTCAACGGCGTCTCGGTTCGGCGGATGCGGACGAAACCCCGTGTTCGGGTCTGGCGACG 3160569
Qy 523 SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 542
Db 3160568 GGACTCGCGCAGGCGCGCGCCAGCGCGGTATGTCGCGCACGGGCTGGGCGGACGAC 3160509
Qy 543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTyr--- 561
Db 3160508 GGCACACGGC-----GCTACTCCACAGCAGCGGGCGGACCTTGGCG 3160467
Qy 562 ---PheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla 580
Db 3160466 CCCTTCGAGGCCAGCCGACATCGCAAGACGACCGGGACCGATCGCCACCAAGTGCC 3160407
Qy 581 AspGlySerArgPheValTyrAla-----ProGlyAspProGlyGlnProVal 596
Db 3160406 GACGGCGGCACACTGCTGTGTGCTCTCGTGCACCTGGGACGGC-----ACG 3160362
Qy 597 ValTyrAla-----ValGlyPheGlyAsnSerTyrAlaAlaAsnSerGlnGlyVal 612
Db 3160361 ACGTACGGCGGCCACCGCTCGACGACACGCGCGAGCTGTGTCGAGGTCTCTCTCTTC 3160302
Qy 613 ProAlaAsnAlaGlnIleArgSerAspArgValAsnProTyrThrPheTyrAla----- 630
Db 3160301 CCGAAGGCGCGCACACCGGTGCGGACCGCGCCGATCGACGCGCTTCTACGCTCAGCAG 3160242
Qy 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAla 650
Db 3160241 TTCGACAATGGAACGCTATACGCCAGCACTGACAGTGCGGTTCGTCACGCCCGGTGCG 3160182
Qy 651 AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlu 670
Db 3160181 GCGGACTGCCCC---TCCGGCAGCAGCAGTTCAGGTGTCGCGGCGCGCGGACGAGC 3160125
Qy 671 GlyAspLeuTyrLeuAlaAlaSer---SerGlyLeuTyrHisSerThrAsnGlyGlySer 689
Db 3160124 GCGGACCTGTGGCTCTCCGCCAAATGGAACGGGCTCTACCGGTCCACCGACGCGGGGAC 3160065
Qy 690 SerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla 709
Db 3160064 ACCTTCGCGAGATCGACAGCTGTGGGCTCTGTCACACCTCTCGGCTTCGGCAAGCGGCC 3160005
Qy 710 ProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla 729
Db 3160004 GACGGCGCGGACTACCGCGCGATCTACCGAGTCGGCTCGCGAGACCATCACCGCCGCTC 3159945
Qy 730 TyrArgSerAspAspCysGlyThrThrTyrValLeuIleAsnAspAspGlnHisGlnTyr 749
Db 3159944 TACCGCTCCGACGACGCGCCGAGCATGGTCCGGATCAACGACGACGCGCCACCATGG 3159885
Qy 750 GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGly 769
Db 3159884 GGGTGGATCGGCGAGGCGGTGCTGCGTGACCGCGCATCCACGCGCGGGTCTACCTCGCC 3159825
Qy 770 ThrAsnGlyArgGlyIleValTyrClyAspIle-----GlyGlyAlaPro 784
Db 3159824 ACCAAGCGGCGCGCATCTCAGTACGGGAGCGCGTGTGATCCCGGGGTCTACGATGCCA 3159765
Qy 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804
Db 3159764 CCGGGCGCGCATCTCTGTTCGTGGCG---ACTACACCCCGAGCAGTGGCGCGGAGG 3159708
Qy 805 Ser-----ProSerSerProSerPro----- 812
Db 3159707 TCTGGCAGGAGGACGTACGCTCTCATGAGGAGGCGCGGTCACTCCGTACCGTCCGCGC 3159648
Qy 813 ---SerProSerProSerSerPro----- 820
Db 3159647 TCTTCTCTGGGCGGAGCTCGAACCCACCGCGGGCGGAGGAGTTCGGGTGGCTCGACC 3159588
Qy 821 -----SerSerSerProSerProSerPro 828
Db 3159587 GGCTCATGGACCTGATGACAGAGAAGCGCATCGGCGTCTGTCGCGCACCCCGCGCT 3159528
```

```
Qy 829 SerProSerPro----- 832
Db 3159527 CGCCCCCGCTTGGATGGGGCGCTTGACCCGAGACCCCTGCCCCGCGACGAGGACGGCC 3159468
Qy 833 -----SerProSerArgSerProSerProSerAla 842
Db 3159467 GCATCGAGTGTGGGGCGCGCCAGCAGCACTTCGCCCACTCCAGTCCGCTTACCGCGCT 3159408
Qy 843 SerProSerProSerSerProSerProSerSer-----ProSerSer 858
Db 3159407 ACGCCGCGCGCATCAACGAGGACCTTCGCCCGCGCTACGGGGGCCATCCCGCCCTCACGA 3159348
Qy 859 SerProSerProThrProSerSerProValSerGlyGlyValLys----- 874
Db 3159347 TGTGGCACATCAACACGAGTACTGCACCTAGCACTGGGGCGGACGAGGGCCACCGCCT 3159288
Qy 875 -----ValGlnTyrIysAsnAspSerAlaProGlyAsp 886
Db 3159287 TCCGCGCGTGGCTCCAGGGCAGGTACGGCAGCTCGACGCACTCAACACGCGCTGG--- 3159231
Qy 887 AsnGlnIleLysPro---GlyLeuGlnValValAsnThrGlySerSerSerValAspLeu 905
Db 3159230 -----GCACGGCGCTTCTGGAGCCAGGCTTACGACACTGGGAAGGCATCTCCCGCCCGC 3159177
Qy 906 SerThrValThrValArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyr 925
Db 3159176 GCCAGCCCACTTACATGAAACACCCACCCAGGCTGCTGCATTCAGCGCTTACCTCCG 3159117
Qy 926 AsnCysAspTyrAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
Db 3159116 ACGCACTCATGGAGTGCT---TCGTCGCGGAAGTACGATCGTCG----- 3159075
Qy 946 ProAlaThrPro 949
Db 3159074 CCGCGCACACCC 3159063
```

```
RESULT 8
US-10-156-761-2561
; Sequence 2561, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2561
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2217)
US-10-156-761-2561
```

```
Alignment Scores:
Pred. No.: 1e-91 Length: 2217
Score: 1442.00 Matches: 328
Percent Similarity: 54.17% Conservative: 107
Best Local Similarity: 40.85% Mismatches: 272
Query Match: 28.08% Indels: 96
```


Db 1903 AATGGAACGGCTCTACCGGTCCACGACGGCGGGAGACACTTTCGCGAGGATCGACAGC 1962
Qy 697 ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAla 716
Db 1963 TGTGGGCTCGTACACCTCGCTTCGCAAGCGCGCGACGGCGCGACTACCCCGCG 2022
Qy 717 ValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGly 736
Db 2023 ATCTACCGAGTGGCTCGACGGAGACCATACCGCGCTCTACCGCTCGACGACGGCGCC 2082
Qy 737 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756
Db 2083 AGGACATGGTCCGGATCAACACGACGCCACACAGTGGGGTGGATCGCGAGCGCGTC 2142
Qy 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776
Db 2143 GTCGGTGACCGCGCATCCACGGCGCGGTCTACCTCGCACCAACGGCGCGGCATCCAG 2202
Qy 777 TyrGlyAsp 779
Db 2203 TACGGGAG 2211

RESULT 9

US-09-927-827-21
; Sequence 21, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; PRIORITY FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIORITY FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 21
; LENGTH: 3668
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2668)
US-09-927-827-21

Alignment Scores:
Score: 3,16e-72 Length: 3668
1167.50 Matches: 299
Percent Similarity: 48.5% Conservatives: 137
Best Local Similarity: 33.30% Mismatches: 351
Query Match: 22.74% Indels: 112
DB: 10 Gaps: 28

US-09-917-376-1 (1-957) x US-09-927-827-21 (1-3668)

Qy 23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42
Db 1136 ACCACACAGGAGCATGGCTGGTGTCTTGGAGCTGTG---CTGTGTCTCTTTTCGAGC 1192
Qy 43 Pro-----AlaHisAlaAlaThrGlnProThrThrTrpSerAsnValAla 58
Db 1193 CCCAGCGTGGCGCGCGAGCCCGCCACGTCCGGCCCTACAGTGGCGGAGTGTGGCC 1252
Qy 59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78
Db 1253 ATTGGCGGTGGCGGCTTGTACCGGTGTGTCTTTCATCCCGCGAGCGTGTGGCC 1312
Qy 79 TyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle 98
Db 1313 TATGGCGCACCGCATGGGTGGCGGTACCGCTGGATGTCGACGCGCGAGCATGGAGC 1372
Qy 99 ProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118

Db 1373 CGCTGACCGAGCTGCTTGGCGCTGACGACCTGGAACCTGATGGGATCGACGATTCGCC 1432
Qy 119 AlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 138
Db 1433 GTGACACCGCGGATCGCGATCGCTGTATCTGGCGCGCGGACCATATATGATGAGAA--- 1489
Qy 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 158
Db 1490 CGCGCGCGCATCTGCCGAGTGTGGCTCGTTCAACCGCGCGCGACAGTTCGAGCGTGC 1549
Qy 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla 178
Db 1550 GACCTGCCGTTTAAAGCTGGTGGTAACACAGCTGGCGCGCGCAATGGCGAGCGGTGGCG 1609
Qy 179 ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198
Db 1610 GTGGACCGCGAGATGGCGCGGTGTCTGTCTGGGTCTCGCGCATGCC---GGCTGTGG 1666
Qy 199 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218
Db 1667 CGTAGCGAGATCGCGCGCGCACTGGGCGAAGTGGCTGCTTTCGCGACGCGCGCTG 1726
Qy 219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238
Db 1727 GCGGTGCCACCGCGCGCAATCATGTGGCGCGAGAGCGGTG---GGGATCGCTTT 1783
Qy 239 ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly 258
Db 1784 GTGTGTTCGACGCGAGCAGTGGCAACAGCGCTGCCAACACCGCGCATCTACGTGGCG 1843
Qy 259 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 278
Db 1844 GTGTCCCGCAACACAGCAGCTGTATGTCTCGAAGATGCCGCGCGAGTTCGGCACCG 1903
Qy 279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
Db 1904 GTGGCGGGCAACACAGCGGCTTGGCGCGCGACCATGGCGCGCGAGCATGGCGGCGCAC 1963
Qy 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318
Db 1964 ---TGTATCTGAGTATGGCGCGACGCGCGCGGAGCTGATGGCGGGGAGCGCTTG 2020
Qy 319 TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338
Db 2021 TCGAAATTCACCGCGCGACAGGGCGCTGGCTGGATGAGATCAGCCCGATTCGCGAG---CCA 2077
Qy 339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
Db 2078 GCCAGTGGCGATGGATTCCGCTGGGTGGGTGGCGGTGGATCCGCAACATCCGCGAGTG 2137
Qy 359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp 378
Db 2138 CTGCTGGCGGACGACCTTCGCGCTCGCACGCGCGCGAGAGCTGTATCGACGCTGGAT 2197
Qy 379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398
Db 2198 GTGGAGACGACTGACGCGCTG-----TTGGCGCAT 2230
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProPro 418
Db 2231 CGGTGTTCGATCACAGCGCGCGCGCTGG---ACCGCACATGCCACGCGCGCAC----- 2281
Qy 419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 438
Db 2282 -----TGGATGGG---GCGCTGGCGATCGATCCGTTCGACGCGC 2317
Qy 439 AspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrp 458
Db 2318 AACCATGCGCTGTCTGTGACCGGCTACCGCATCTGGGCTTCGCGCAATCTG----- 2368
Qy 459 AspSerGlyGlnIleHisIleAlaPro-----MetVal 470
Db 2369 -----CAGATTCGCGCGCACCGCGCGCTGCGTGGTGGTTCAGGAC 2416

Db 393 ATCGCGCTGGACCCCAACACCCGACAGGCTGTACCTCGCGCAGGGCGCTATGTCCGC 452
Qy 136 AsnSerTrpAspProAsnAspGlyAlaIleuArgSerSerAspGlnGlyAlaThrTrp 155
Db 453 GACGAGTGG-----GCGCGCTTCTATGTGTCGGAAGACCCGCGCCAGCTCGTTT 500
Qy 156 GlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyValGlyMetGlyGlu 175
Db 501 ACAATCTACAGATCGCGCTTCCGATGGCGCCCAACGACATGGAGCGCAATATGGCGAG 560
Qy 176 ArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyPheGlyAlaProSerGlyLys 195
Db 561 CGCTCGCTGTCAACCCGTTTCAACTCGAAGCAGGCTCTGGATGGGTACGCGTACA---GAG 617
Qy 196 GlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAsp 215
Db 618 GGTATCTGGAAGATGTGGACCGCGCCCAAGACCTTGGACAAACGTCACGTCATCCCGGAC 677
Qy 216 ValGlyThrTyrlleAlaAsnProThrAspThrThrGlyTyrlleGlnSerAspIleGlnGly 235
Db 678 GCGTTC-----ACCAACGGTATCGGATACAGTGC----- 707
Qy 236 ValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIle 255
Db 708 -----GTCAATTTTCGACCCC-----GAAAGTAAATGGCACCATC 740
Qy 256 PheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThr 275
Db 741 TACGCGAGCGCGACTCGCCCGCACGGC---ATGTACGTACGCGACGACGCGGTGTCTCG 797
Qy 276 TrpGlnAlaValProGlyAlaProThrGlyPheIle----- 287
Db 798 TGGAGCCAGTGGCGGCGCCAGCCGTCCAGCTGGGTCAAACAGACCAACGGCGGTTCCCG 857
Qy 288 -----ProHisLysGlyValPheAspProValAsn 297
Db 858 GACAAGAGCCGCGTGCATCGCGCGCGCCATGAAGTCGTCTCACCCCC----- 911
Qy 298 HisValLeuTyrlleAlaThrSerAsnThrGlyGlyProTyPheAspGlySerSerGlyAsp 317
Db 912 AACTTCTCTACGTGACTTACGCGGACTTACCTGCTGCTCCATGGCGGCGTCACGTTGGCGAA 971
Qy 318 ValTrpLysPheSerValThrSerGlyThrTrpThrArgIle----- 331
Db 972 GTCTGGCGCCAGAACCGCAGCTCGGGCGCCCTGGGACGACATCTCCCGCGTGGGCAAC 1031
Qy 332 ---SerProValProSerThrAspThrAlaAsnAspTyPheGlyTyPheSerGlyLeuThr 350
Db 1032 TCGTGGCTGCGCGTACAAACACGAGCGTTCCTCGCGGCGGATTTGCGGTTCTCAGC 1091
Qy 351 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 370
Db 1092 GTCAGCGGACCAACCCCAACCGTCTCGTGTCTATCACC---CTCGACCGGACCCCGGA 1148
Qy 371 ThrIle-----IlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrp 387
Db 1149 CCCGCCCTCGACGATCTACCTCTCAACCGATCGCGCGGACCTGGAAGACGTCACC 1208
Qy 388 AspTrpThrSer-----TyrProAsnArgSerLeuArg 398
Db 1209 CAGTCTGCTCCCGTCCAACTCGAAGGTAACTGGGGCCACCCGACTAAACGCGCGCGG 1268
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPro----- 415
Db 1269 TAC---AAGAGCGCACGCGCTTTCGTTGGCTCGACTTCAACACCGTCCCGCAGTGGGG 1325
Qy 416 -----AsnProValProSerPro-----LysLeuGlyTyPheMetAspGluAla 430
Db 1326 GGATACGCTGGCGCGCAGCGTACGCGCGCCCTCACCAAGTTTGGCTGTGGATGACGCGCT 1385
Qy 431 MetAlaIleAspProPheAsnSerAspArgMetLeuTyrlleGlyAlaThrLeuTyrlle 450
Db 1386 GTGCTTATCGATCCGTTCAACCCCGACGACCTGTATGGCAGCGGGGCGACCATCTGG 1445

Qy 451 AlaThrAsnAspLeuThrLys-----TrpAspSerGlyGlyGlnIleHisIle 466
Db 1446 GCGACCGACACGCTCTCCCGTGTCCAGAAAGACTGG----- 1481
Qy 467 AlaPro-----MetValLysGlyLeuGluGluThrAlaValAsnAspLeuIle 482
Db 1482 GCGCGGAGCTGGTACCTCCAGATCCAGGTATCGAGGAGATGCGATCTGTGCTCGCTCCG 1541
Qy 483 SerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 502
Db 1542 TCGCCCAAGAGCGCGCGCTCTCTCGGCGCATCGGTGACATTAGCGCATGAAGCAC 1601
Qy 503 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr 522
Db 1602 GAGGACCTCAACCAAG---CCCAGAGATGTTTGGTGGCCCCAGTCTTCCACCTCGAC 1658
Qy 523 SerValAspTyPheAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 542
Db 1659 AGCATCGACGCTCGCGGCAACTTCCCAACGTTGTCTCCGCGCGGATCTCTCGGACAC 1718
Qy 543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp--- 561
Db 1719 GAGTACGACAGCGCTGCGCGCGCTGCGTACGCGACTGACGCGGAGACGCGTGGACC 1778
Qy 562 ---PheGlnGlySerGluProGlyGlyValThrThr-----GlyGlyThrValAla 577
Db 1779 ATCTTCCCTACTGCTCCCTCTGCGCATGAACGCGAGCACCTACCAGGCGCACGATGCA 1838
Qy 578 AlaSerAlaAspGlySerArgPheValTrpAlaPro-----GlyAspProGlyGlnPro 595
Db 1839 GTCGACGCGAGCGCGCAGCATCTGTGTGTCGCAACCAAGCTTACGACGAGGCGCTCGGA 1898
Qy 596 ValValTyrlleAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla--- 614
Db 1899 CCGTGTACTCGCACGACTATGCGCAAGCGTGTCT-----GTTCCGCTGGC 1946
Qy 615 -----AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrlle 629
Db 1947 GACCTGAAGCCCGACGACTGCCAATGTCTCTCGACAGGTCCAGATGCGACGTTCTAC 2006
Qy 630 AlaLeuSerAsnGlyThrPheTyPheArgSerThrAspGlyGlyValThrPheGlnProVal 649
Db 2007 GCTACCGATGCGCGCAAGTCTTCTGTCGACGCGCGGGAAGTCGTATGCGGCCAAG 2066
Qy 650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 669
Db 2067 GCGCGCGGACTTGTCACT-----GSCACATCGCTCATGCTCGCGTGAACCCCTGGGTG 2120
Qy 670 GluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyPheHisSerThrAsnGlyGly 688
Db 2121 GCGCGCGACGCTGCGGTGCTGTCCCGAGGCGGTCTCTTCCACTCGACCGACTTTGGC 2180
Qy 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707
Db 2181 GCTGTTCACGAGGTAGTACCCGCAACGCGACCTCTGTGAGCGTCCGCGGCCCAAG 2240
Qy 708 Ser-----AlaProGlySerSerTyPheProAlaValPheValValGlyThr--- 722
Db 2241 TCCAAGTCGACGCGCAAGAGCTAGCGCGCTCCGCGCTCTTCATCTCGGCGCACCGAC 2300
Qy 723 ---IleGlyGlyValThrGlyAlaTyPheArgSerAspAspCysGlyThrTrpValLeu 741
Db 2301 AAGCCTGGAAGCGACATCGCGCTGTACCGCTCCGACGACCAACGCGACACTGACCGCGC 2360
Qy 742 IleAsnAspAspGlnHisGlnTyPheGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761
Db 2361 GTCAATGACGAGGACCACTACTCTCGGC---CCACCATGATCGAGGCGGCCCCAAG 2417
Qy 762 AsnLeuArgValTyrlleGlyThrAsnGlyArgGlyIleValTyrlleGlyAspIle 780
Db 2418 GTCTACGCGCGCTGTATCTAGGCACGACCGCGCGGTATCGTGTACGCGGACCTT 2474

RESULT 11
US-10-395-241-17
; Sequence 17, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Xyloglycan Oligosaccharide-Degrading Enzyme with Histidine Tag
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
US-10-395-241-17

Alignment Scores:
Pred. No.: 2,38e-69 Length: 2481
Score: 1124.00 Matches: 298
Percent Similarity: 48.91% Conservative: 127
Best Local Similarity: 34.29% Mismatches: 306
Query Match: 21.89% Indels: 138
DB: 18 Gaps: 33

US-09-917-376-1 (1-957) x US-10-395-241-17 (1-2481)

QY	52	TyrThrTrpSerAsnValAlaIleGlyGlyPheValAspGlyLeuValPheAsn	71
DB	13	TACGAGTTCACAGATGTCGATCGCGCGCGGATACATACCGGGATTCGCGCAC	72
QY	72	GlulGlyAlaProGlyLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp	91
DB	73	CCAAAGACCAAGGACCTGCTGACGCGCGCACGATTCGGCGCGCTACCGCTGGGAC	132
QY	92	AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	111
DB	133	GCAGGCGATCCAGTGGATCCGCTCAACGACTTATCGAGCGCGACATGACATT	192
QY	112	AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal	131
DB	193	ATGGGCGCGAGTCGATCGCTGGACCCCAACACCCCGACAGGCTGTACCTCGCGCAG	252
QY	132	GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleuArgSerSerAsp	150
DB	253	GGCGCGTATGTCGCGCACAGTGG-----GCGCGCTTCTATGTGTCGGAAGAC	300
QY	151	GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly	170
DB	301	CGCGCGCGATGCTTACATCTACGAGTCCGCTCCGATCGGCGGCCAACACATGGGA	360
QY	171	ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleuTyrPheGly	190
DB	361	CGCAACAATGGCGAGCGCTCGCTGTCAACCCGTTCAACTCGAAGTCGAGTCTGGATGGGT	420
QY	191	AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet	210
DB	421	ACGCGGTACA---GAGGATATCTGGAGAGTTCGGACCGCGCCCAAGACCTGGCAACAGTC	477
QY	211	ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln	230
DB	478	ACGTCCATCCCGACCGCTTC-----ACCAACGGTATCGGATACAG	519
QY	231	SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln	250
DB	520	TCG-----GTCAATTTTCGACCCC-----GAA	540
QY	251	AlaSerLysThrIlePheValGlyValAlaAAspProAsnAsnProValPheTrpSerArg	270
DB	541	CGTATATGGCACCATTCTACCGAGCGGACTGCCCCCGCAGGCG---ATGTACGTACACGAC	597
QY	271	AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle-----	287
DB	598	GACGCGGTGTCTCGTGGAGCAGGTGGCGGCGCAGCTCCAGCTGGCTCAACAGGAC	657
QY	288	-----ProHisLysGlyVal	292
DB	658	ACGGCGCGTTCCTCCGACAAAGCCCGCTCGATCGCGCGCAGCCCATGAAGTCGCT	717
QY	293	PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp	312
DB	718	CTCACCCCC---AATCTCTTACTAGTACTTACGCGGACTACCTGGTGTCCATGGGC	771
QY	313	GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle---	331
DB	772	GTACGTTTCGCAAAATCTGGCGCCAGAACCGCACCTCGGCGCGCTCGGAGGACATTACT	831
QY	332	-----SerProValProSerThrAspThrAlaAsnAspTyrPheGly	345
DB	832	CCCCCGCTCGGCAACTCGTCGCTGCCCGGTACACACACAGACGTTCCCTCGGCGGGA	891
QY	346	TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle	365
DB	892	TTTTTGGGTCTCAGCTCGAGCGGACCAACCCCAACCGTCTCGTCTCATCACC---CTC	948
QY	366	SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr	382
DB	949	GACCGCGACCCCGGACCCGCTCGACAGCATCTACTCTCAACCGATCGGCGCGGAC	1008
QY	383	TrpThrArgIleTrpAspTrpThrSer-----TyrPro	393
DB	1009	TGGAAGGACGTACCCAGCTCTCGTCCCGTCCACCTCAAGTAATACTGGGCGCACCG	1068
QY	394	AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal	413
DB	1069	ACTAACCGCGCGGTAC---AAGGACGCGACGCTCTCTCGTGGCTCGACTTCAACAC	1125
QY	414	GlnPro-----AsnProProValProSerPro-----LysLeuGly	425
DB	1126	GGTCCCGACGTTGGGGGATACGCTGGCGCGCATCGGTACGCGCGGCTCACCAGTTTGGC	1185
QY	426	TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr	445
DB	1186	TGTTGGATGAGCGCTGTGCTTATCGATCCGTTCAACCCCGACGACCTGTATGTACGGCAG	1245
QY	446	GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly	461
DB	1246	GGGCGCACCATCTGGGCGACCGACACGCTCTCCGCTCGAGAGGACTGG-----	1296
QY	462	GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla	477
DB	1297	-----GCGCGGAGTGGTACCTCCAGATCGACGCTATCGAGGAGATGCG	1341
QY	478	ValAsnAspLeuIleSerProSerGlyAlaProIleLeuSerAlaLeuGlyAspLeu	497
DB	1342	ATCTCTGCTCGCTCGCGCCAAAGGCGCGCGCTCTCTGTCTCGGCATCGGTGACATT	1401
QY	498	GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal	517
DB	1402	ACGGGATGAGCAGCAGCAGCTCACCAAG---CCCCAGAGATGTTGGTGGCGCCCGAC	1458
QY	518	PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla	537
DB	1459	TTTCCCAACCTCGACAGCATCGACGCTCGGCGCAACTTCCCAACGTTGTCTCGCGCC	1518
QY	538	GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly	557

Alignment Scores:

Pred. No.: 4,68e-56 Length: 5698
 Score: 940.00 Matches: 374
 Percent Similarity: 30.11% Conservative: 136
 Best Local Similarity: 22.08% Mismatches: 315
 Query Match: 18.31% Indels: 871
 DB: 18 Gaps: 59

US-09-917-376-1 (1-957) x US-10-420-191-1 (1-5698)

Qy 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
 |||||
 Db 128 TCTCAGTCTTCTTCCAGAGAGCTCAGGAACGGGAACAGGACCCCGCGAGTAGGACGG 169
 |||||
 Qy 45 HisAlaIleThrGln----- 50
 |||||
 Db 170 CATGCTGCTTCTTCCAGAGAGCTCAGGAACGGGAACAGGACCCCGCGAGTAGGACGG 229
 |||||
 Qy 51 -----ProTyrThrTrpSerAsnValAlaIle--GlyGlyGlyGlyPheValAspGly 67
 |||||
 Db 230 GTACGACGGAAATTCATGAAGAAGCTCAAGCTCGGCGCGCGCGCTTCGTCCCGGCG 289
 |||||
 Qy 68 IleValPheAsn-----GluGlyAlaPro----- 75
 |||||
 Db 290 ATCATCTTCCAAAGTACCTTCTTCAGTTCAGCGCGCGCGCGCGCGAGGAGCGGCG 349
 |||||
 Qy 76 -----GlyIleLeuTyrValArgThrAspIleGlyGlyMet 87
 |||||
 Db 350 TAGTAGAAGGTTCCCAAGACAAAGCGCTAGCATATGCACGAACAGATATTGGCGGCTG 409
 |||||
 Qy 88 TyrArgTrp-----AsnAsnTrpGly----- 90
 |||||
 Db 410 TACCGCTTCAAAGGTTCTGTTTCCGATCGTATACGTGCTGTGTCTATACCGCGCGAC 469
 |||||
 Qy 91 -----AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal----- 104
 |||||
 Db 470 ATGGCGGAGTTCGCGAGCATCTATGACCGCGCGCGCGCGCGAGTGGATTCGTATATGCC 529
 |||||
 Qy 105 GlyTrp----- 110
 |||||
 Db 530 GGCTGGCAACAGCGGCTGCTGAGTACCTTGGCGGCGAGTGCCTACCTAACGACTATTACGG 589
 |||||
 Qy 111 -----TyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal 127
 |||||
 Db 590 CCGACCGGTGTTCTGGGGCATCGACGCTGTGGCTTGATCCGCGAGGACATCAAAAGGTG 649
 |||||
 Qy 128 TrpAlaAlaVal----- 131
 |||||
 Db 650 TATGCCGAGTGATACCCGCTAGTGGCAACACCGGNACTAGGGGCTCTAGTATTTCAC 709
 |||||
 Qy 132 -----GlyMetTyrThrAsnSerTrpAsp----- 139
 |||||
 Db 710 ATACGGCTCACGGCATGTATACGAACAGCTGGTC-TGTGATGTCTCTCAGATCTAGACC 768
 |||||
 Qy 139 ----- 139
 |||||
 Db 769 TATGATGGACGGCGGTATATATGCTTGTTCGACCAACACTACAGAGAGTCTAGATCTGG 828
 |||||
 Qy 140 -----ProAsnAspGlyValAlaIleLeuArg 147
 |||||
 Db 829 ATACTAAGCTGCTGACATGCGCCATATATAGGGATCCGAGTATGGAGCCATTCATTCCG 888
 |||||
 Qy 148 SerSerAspGln----- 151
 |||||
 Db 889 TCGTCAGACCGGACTGTAAACGGGTATATATCCCTAGGCTCATTTACCTCGGTAGTAAGCG 948
 |||||
 Qy 152 -----GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 167
 |||||
 Db 949 AGCAGCTGGCGGGCGCAACGTGGTCTTCCAACTTGGCCCTTCAAAAGTCGGGGTAAC 1008
 |||||
 Qy 168 MetProGlyArg----- 171
 |||||

Db 1009 ATGCCAGGACGCCCGCGTTCACACCGAAGTGTTCACACCGGAAGTTCAGCCCCCATTTG 1068
 |||||
 Qy 172 -----GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 187
 |||||
 Db 1069 TACGCTCTCGGGAGCCGGAGAGCGTCTGGCTGTGATCCCGCAACTCCACATCATC 1128
 |||||
 Qy 188 TyrPheGlyAlaPro----- 192
 |||||
 Db 1129 TACTTTGGTCTCTCTCGGCCTCTCGCAGACCGACAGTAGCGCGTGTAGGTGTAGTAG 1188
 |||||
 Qy 193 -----SerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrp 207
 |||||
 Db 1189 ATGAACACACGACGCTCAGAAACCGCCTCTGGAAGTCTACGAGCGCGCGTACCTTT 1248
 |||||
 Qy 208 SerGlnMet----- 210
 |||||
 Db 1249 TCCAAGGTCTCGCGAGTCTTTTTCGGGAGACCTTCAGATGCTGCCCGCGACTCGAAA 1308
 |||||
 Qy 211 -----ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrTrp 227
 |||||
 Db 1309 AGGTTCACAGAGCTCGTTTACGGCAACTGGGACGTACATCCAGACCCGAGTGATTCACA 1368
 |||||
 Qy 228 GlyTyrGln----- 230
 |||||
 Db 1369 GGCTACAAACAGCAGCAGCAAGTGCCTGACCCCTGCATGTAGGGTCTGGGGTCTACTAAGTTG 1428
 |||||
 Qy 231 -----SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 247
 |||||
 Db 1429 CCGATGTTGTTCGACAAAGGACTCATGTGGTTACGTTTCGACTCAACACGACGACG 1488
 |||||
 Qy 248 LeuGlyGlnAla----- 251
 |||||
 Db 1489 ACCGGGAGCGCTGTTCTCTGAGTACACCAATGCAAGCTGAGTTGGTCTGCTGTCG 1548
 |||||
 Qy 252 -----SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe 267
 |||||
 Db 1549 TGGCCCCCTCGGACGCTCTCGTATCTTGTGTGGCAGCGCTGATAACATCATCTGCTTCAGTC 1608
 |||||
 Qy 268 Trp----- 268
 |||||
 Db 1609 TATGTGACACGTGCAGAGCATAGAAAACACCGTCCGACTATTGTAGTGACGAATCAG 1668
 |||||
 Qy 269 -----SerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286
 |||||
 Db 1669 ATACACTCGTCAATGCCGCTCCAGCTGAGTGTGTACCGGGGAGCCAGCCAGGAATAC 1728
 |||||
 Qy 287 IleProHis----- 289
 |||||
 Db 1729 TTTCTCTCAAGTTACGGCGGAGGTGCACCTCAGCAGATGGCCCGCTCGCTTCTTATG 1788
 |||||
 Qy 290 LysGlyValPhe-----AspProValAsnHisValLeuTyrIleAlaThrSerAsn 306
 |||||
 Db 1789 AAAGAGGTGTTCCGGAACCTGCAGCCAGCAGAGAGGCTTGTATCTGACCTATTCCGAT 1848
 |||||
 Qy 307 ThrGlyGlyPro----- 310
 |||||
 Db 1849 GGCAAGGCGCGCGCTTTGACGTCGTCTCTTCGGAACATAGACTGGATAAGGCTA 1908
 |||||
 Qy 311 -----TyrAspGlySerGlyAspValTrpLysPheSerValThrSerGly 326
 |||||
 Db 1909 CGGTGTCGGGCTATGATGGCACACTTGGCTCAGTGTGGAGGTACGACATTCGAGGGGA 1968
 |||||
 Qy 327 ThrThrThrArgIleSerProValProSerThrAspThr----- 339
 |||||
 Db 1969 ACTTGGAAAGACATACCTACCGTGTGAACCGAGTCAACCTCCATGCTGTAAACGTCGCCCT 2028
 |||||
 Qy 340 -----AlaAsnAspTyrPheGlyTyrSerGlyLeu 349
 |||||
 Db 2029 TGAACCTTTCTGATCACCCCTGTCTCTGGATCAGATCTATCTTTGGCTTTGGCGCCT 2088
 |||||
 Qy 350 ThrIleAsp----- 352
 |||||
 Db 2089 GGCCTCGATTGTAGTGGGGACAGAGACCTAGTCTAGATATGAACCGGAACCGCCGGA 2148
 |||||


```
QY 649 -----ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666
Db 979 CGGTCGTGGCGCGTGGCTTCCAGCAGC-----ATCGGCATGCTCATCGACACCTTA 1032
QY 667 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 686
Db 1033 -----CGCAAC 1038
QY 687 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal----- 703
Db 1039 GGTGGGTGGTGGCAACGACCAACAGCCGCGAGCAGCCGCGAGCGATGTCAACACCTTC 1098
QY 704 -----Gly 704
Db 1099 GTCAACAGTCGAGATTGACTCTCGGCGAGCAGCCGCGCTGTGGTCAACAGAACGGT 1158
QY 705 PheGlyLysSerAlaProGlySerSerTyrProAlaValPhe----- 718
Db 1159 GCGGCGCTCGCCAGCCGCGCAGCAAGCCGCGAGCTTCCCGAACGCGCACCTCGAC 1218
QY 719 -----ValValGlyThrIleGlyValThrGlyAla----- 729
Db 1219 CGGTATGTCGTGATCAAGCCGCGGTGAGTCGCGAGCGCACAGCGCTCGCGAGCGATCCG 1278
QY 730 -----TyrArgSerAspAsp---CysGlyThrThrTrpValLeuLeuAsnAspAsp 745
Db 1279 ACAACTGCAAGAAGTGGACCCCATGTGCGACCGCGATACAGCAGC----- 1326
QY 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
Db 1327 -----TCGTACGGGGTACTGACCAACCGCTTACCGAACTCCCGCATCGCGCCAGTGG 1380
QY 761 AlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 1381 TTCCCGCGCAGTTTGACCGCTTGTGCGAAGCAGCGCCAGCGGTGCCGAGCTCGACC 1440
QY 781 GlyGly-----AlaProSerGlySerProSerProSerValSer 793
Db 1441 AGCTCGAGCCCGCGCTTCCGCGCGAGTCGCTCGGCTTCGCGAGTCCGAGCCCGAGT 1500
QY 794 ProSerAlaSerProSerProSerProSerProSerProSerProSerProSer 813
Db 1501 CCGAGCCGCGAGCGTCCGCTCGCGCTTCCGAGCTCGAGCCCGCTCGCGTCCGCTCG 1560
QY 814 Pro-----SerProSerSerProSerProSerProSerProSer 825
Db 1561 CCGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGTCCGCTCTCCGAGCTCGAGCCCGTCT 1620
QY 826 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 845
Db 1621 CCGTCCGCGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGTCCGCTCTCCGAGCTCGAGC 1680
QY 846 ProSerSerProSerProSerProSerProSerProSerProSerProSerProSer 865
Db 1681 CCGTCTCGCTCGCGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGTCCGCTCTCCGAGC 1740
QY 866 SerSerProValSerGlyGlyValGlnTyrLysAsnAsnAspSerAlaProGly 885
Db 1741 TCGTCGCGGTGTGGGTGGGCTGAAGGTGCGAGTACAGACAAATGATTCGCGCGCGGT 1800
QY 886 AspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerSerValAspLeu 905
Db 1801 GATAACCAAGATCAACACCGGGTCTCCAGTTGGTGAATACCGGGTCTCGTGGTGGATTG 1860
QY 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925
Db 1861 TCGACGGTGACGTGCGGTACTGGTTTACCCCGGGATGGTGGGTCTCGCACTGGGTGAC 1920
QY 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
Db 1921 AACTGTGACTGGGCGCGCATGGGGTGTGGGAATATCCGGCGCTCTGTTGGCTCGGTGAAC 1980
```

```
QY 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 1981 CGGCGGACGCGCGACGCGGACACCTACTCTGCAG 2013
RESULT 15
US-09-917-383-2
; Sequence 2, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3687
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-2
```

Alignment Scores:

Pred. No.:	4,86e-48	Length:	3687
Score:	823.00	Matches:	251
Percent Similarity:	47.07%	Conservative:	46
Best Local Similarity:	39.78%	Mismatches:	173
Query Match:	16.03%	Indels:	164
DB:	10	Gaps:	24

US-09-917-376-1 (1-957) x US-09-917-383-2 (1-3687)

```
QY 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 307 TGGATGGACCGCATCGCTCGCATCAACGGCGTCAACGGC-----GGACCC 351
QY 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 465
Db 352 GCGTTGACGACATATCTGGACCGCGCCCTCTCCAG-----CAGCAGGA 396
QY 466 IleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuLeuSerProPro 485
Db 397 ACCACCCCTGAAGTC-----ATTGAGATTGTCACTACGATCTG-----CCGGGA 441
QY 486 SerGlyAlaProLeuLeuSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspVal 505
Db 442 CCGGACTGCGCGCGCTCGCTTCCCAACGGCGAAGTGCCTGCT---ACGGCAGCAGGTTTG 498
QY 506 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 525
Db 499 CAGACCTAT---GAAACGCGAGTACATCGATTCGAGTATCTCTGAGCAATCCGAAG 555
QY 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerGln 545
Db 556 TACTCCAGCTCGGATCGGTGACGATCAT---GAGCCGAGTCTCGT 600
QY 546 ProAsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySer 565
Db 601 CCNAACGC---GGTCAC-----CAATAT 620
QY 566 GluProGlyGlyValThrThrGlyThrValAla-----AlaSerAlaAspGly 582
Db 621 GAGCATTCAGCGTGTGCAACGCGGTGCGGTATT-ACGAGCAAGGATCGAGTACGCGC 679
QY 583 SerArg-----PheValTrpAlaProGlyAspPro--- 592
```


	1681	CGGTCTCCGTCGCCAGACCCGAGTCCGAGCCCGAGTAGCTCGCGCTCGCGTCTCCGACG	1740
Db			
	866	SerSerProValSerGlyGlyValVysValGlnTrpYrsAsnAenAepSerAlaProGly	885
Qy		:::::	
	1741	TCTGTCGGCGGTGTCGGTGCGGTGAAGTGCAGTACAAGAACAATGATTTCGGCGCCGGGT	1800
Db			
	886	AspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerValAspLeu	905
Qy		:::::	
	1801	GATAAACAGATCAAAACCGGGTCTCCAGTTGGTGAATACC GG GTCTGTCGTTGGATTG	1860
Db			
	906	SerThrValThrValArgTrpTrpPheThrArgAspGlyGlySerSerThrLeuValTrp	925
Qy			
	1861	TCGACCGTGCAGCGTGCGGTACTGGTTCACCCGGGATGGTGGGTCTGCACACTGGTGTAC	1920
Db			
	926	AsnCysAspTrpAlaAlaIleGlyCysGlyValAsnIleArgAlaSerPheGlySerValAsn	945
Qy		:::::	
	1921	AACGTGACTGGCGGGCATGGGTGTGGGAATATCCGCCCTCTGTTCCGCTCGGTGAAC	1980
Db			
	946	ProAlaThrProThrAlaAspThrTrpLeuGln	956
Qy			
	1981	CCGCGCAGCGCGACGGCGGACACTACTCTGCAG	2013
bB			

Search completed: October 6, 2005, 01:57:00
Job time : 14605.5 secs

Db	680	TCACGAAATTGCACGCCATTTCGACAGCTGTACATCTACATGGACGCGGCCCATCTCGGCT	739
Qy	593	-----GlyClnProValVal-----TyrAlaValGlyPheGlyAsnSerTrp	606
Db	740	GGCTTGCGTGGCCCAATAATATATCCGACGCGGATACGTACAGGAAGTCCAGAAAGTCTTCAACG	799
Qy	607	AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAen-----	624
Db	800	CGAGCATCGGGGTCAACGGCATCGAGGCTTCGTCCACCAACACGCGGAATTACACGCCGT	859
Qy	625	-----ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr	637
Db	860	TGAAGGAGCGGTTTCATGACCGCCACCCAGCAGGTCGCGCAGACGCGGTGGAGTCGCGCA	919
Qy	638	ArgSerThrAspGlyValThrPheGlnPro-----	648
Db	920	ATTTCCTACCATGGAAATTCGATC- GACGAAGCCGACTACCGGTTGATTTGTACTCG	978
Qy	649	-----ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal	666
Db	979	CGGCTGTCGCGCTGCTTTCACGAGC-----ATCGCATGTCTCATCGACCTTCA	1032
Qy	667	ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn	686
Db	1033	-----CGCAAC	1038
Qy	687	GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal-----	703
Db	1039	GGTTGGGGTGGTCCGAAACGAACCAACAGGCCCGACACCCGACCGGATGTCAACACCTTC	1098
Qy	704	-----Gly	704
Db	1099	GTCACACAGTCGAAGATTGACCTTCGGCAGCACCGCGGCTGTGGTCCACACAGACGGT	1158
Qy	705	PheGlyLysSerAlaProGlySerSerTyrProAlaValPhe-----	718
Db	1159	GCGGCGCTCGGCCAGCGCCGACGAGCAACAGCCGACGACCTCCGGAACGCGCAGCTCGAC	1218
Qy	719	-----ValValGlyThrIleGlyValThrGlyAla-----	729
Db	1219	CGGTATGCTCGATCAACCGCGCGGTGATCGGACCGCAACAGCTGCGGAGCATCGC	1278
Qy	730	-----TyrArgSerAspAsp-----CysGlyThrTrpValLeuIleAsnAspAsp	745
Db	1279	ACAACTGGCAAGAGTCGACCCCATGTGGACCCGACGTACACGAG-----	1326
Qy	746	GlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis	760
Db	1327	-----TCGTACGGGGTACTGACCAACGCGTTACCGAACTCCCGGATCGCGGCCAGTGG	1380
Qy	761	AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle	780
Db	1381	TTCCCGCGCGCATTTTGACCACTGTTCGCGAAGACGACGCGCGGTGCGGACGTGCAAC	1440
Qy	781	GlyGly-----AlaProSerGlySerProSerProSerValSer	793
Db	1441	AGCTCGAGCCCGCGCTCCGCGCCGAGTCCGTGCGTTCGCGAGTCCGAGCCCGAGT	1500
Qy	794	ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer	813
Db	1501	CCGAGCCGACAGCTCGCCATCGCGTCCGCGTCTCCGAGTCCGAGCTCGAGCTCGCTCG	1560
Qy	814	Pro-----SerProSerSerSerProSerSerProSer	825
Db	1561	CCGAGCCCGAGTCCGAGCCCGAGTGTGCGCGTCTCCGAGTCCGAGCTCGAGCTCGTCT	1620
Qy	826	ProSerProSerProSerProSerProSerArgSerProSerProSerAlaSerProSer	845
Db	1621	CGTCCGCGAGCCCGAGTCCGAGCCCGAGTAGTCTCGCGTCCGCTCTCCGAGTCCGAGC	1680
Qy	846	ProSerSerProSerProSerSerProSerSerProSerProSerProSerProSer	865

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:48:48 ; Search time 141.761 Seconds
(without alignments)
2018.911 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ARTQPYTWSNVAIGGGGFVD.....YIGTNGRGIVYGDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4036	100.0	740	6	Abp73016 Amino aci
2	4036	100.0	740	7	Add22926 Acidother
3	4036	100.0	740	7	Add22923 Acidother
4	4036	100.0	957	6	Abp73015 Amino aci
5	4036	100.0	957	7	Add22921 Acidother
6	4036	100.0	957	8	Ado52313 Acidother
7	3920.5	97.1	733	8	Ado52315 A. cellu
8	2063	51.1	940	6	Abp99489 Amino aci
9	1680	41.6	726	7	Add22927 Aspergill
10	1579	39.1	838	7	Add42055 Trichoder
11	1557.5	38.6	818	8	Adh51581 Trichoder
12	1117	27.7	789	8	Adh19120 Geotrichu
13	1117	27.7	812	8	Adh19118 Geotrichu
14	1114	27.6	826	8	Adh19124 Geotrichu
15	979	24.3	756	8	Adr90298 Geotrichu
16	979	24.3	757	8	Adr90302 Geotrichu
17	979	24.3	776	8	Adr90296 Geotrichu
18	841.5	20.8	555	7	Add44919 Xanthomon
19	238	5.9	2468	6	Abu38411 Protein e
20	238	5.9	2468	6	Abp59933 Microbia
21	219.5	5.4	2736	7	Abu81481 Pseudomon
22	199.5	4.9	2435	6	Abu19529 Protein e
23	193	4.8	1465	8	Adr24922 Bacterial
24	181.5	4.5	3892	6	Ada34216 Acinetoba
25	181	4.5	1308	6	Abu16642 Protein e

ALIGNMENTS

RESULT 1
ABP73016

ID ABP73016 standard; peptide; 740 AA.

XX AC

XX ABP73016;

XX 03-JUN-2003 (first entry)

DE Amino acid sequence of the avicelase Aviii catalytic domain.

XX Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
KW cellulose.

XX Acidothermus cellulolyticus.

PN WO2003012090-A2.

XX PD 13-FEB-2003.

XX PF 28-JUL-2001; 2001WO-US023818.

XX PR 28-JUL-2001; 2001WO-US023818.

XX (WIDE) MIDWEST RES INST.

PI Ding S, Adney WS, Vinzant TB, Himmel MB;

DR WPI; 2003-248177/24.

XX New thermostable Aviii peptide from Acidothermus cellulolyticus, useful
PT for degradation of cellulose or in generating anti-Aviii antibodies for
PT purifying recombinant Aviii polypeptides from genetically engineered
PT host cells.

XX Claim 6; Page 8; 44pp; English.

XX The present sequence is derived from a thermostable avicelase, designated
CC Aviii. Aviii is a member of the glycoside hydrolase family of enzymes,
CC and is a cellulase. Aviii is useful in the conversion of biomass to
CC biofuels and biofuel additives. It may be useful in the production of
CC detergents, pulp and paper processing, food and feed processing and in
CC textile processes. The thermostable Aviii peptide is useful in the
CC degradation of cellulose, and in generating recombinant anti-Aviii
CC antibodies that are useful in purifying recombinant Aviii polypeptides
CC from genetically engineered host cells, in detecting Aviii polypeptide
CC expression, as well as a reagent tool for characterizing the molecular
CC actions of the polypeptide. The Aviii polynucleotide is useful as a

Db 421 APWKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY 480
Qy 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKMFQGSERPGVTTGGTVAASADGSRFV 540
Db 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKMFQGSERPGVTTGGTVAASADGSRFV 540
Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDDGGVTF 600
Db 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDDGGVTF 600
Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
Db 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
Qy 661 KSAPGSSYPVAVVGTGGVTVGAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRRV 720
Db 661 KSAPGSSYPVAVVGTGGVTVGAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRRV 720
Qy 721 YIGTNGRGIVYGDIGGAPSG 740
Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 3

ADD22923 ID ADD22923 standard; protein; 740 AA.
XX AC ADD22923;

XX DT 15-JAN-2004 (first entry)
XX DE Acidothermus cellulolyticus avicelase AvIII catalytic domain.

XX KW enzyme; AvIII; cellulose reduction; agricultural biomass;
XX KW municipal solid waste; glycoside hydrolase; avicelase.
XX OS Acidothermus cellulolyticus.

XX PN US2003108988-A1.
XX PD 12-JUN-2003.

XX PF 18-OCT-2002; 2002US-00155400.
XX PR 28-JUL-2001; 2001US-00917376.

XX PA (DING/) DING S.
XX PA (ADNEY/) ADNEY W S.
XX PA (VINZ/) VINZANT T B.
XX PA (HIMM/) HIMMEL M E.

XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX DR WPI; 2003-810853/76.

XX PT New isolated thermal tolerant avicelase polynucleotide useful for
XX PT detection of a polynucleotide encoding AvIII and for reducing cellulose
XX PT in a starting material, e.g. municipal solid waste.

XX PS Claim 16; SEQ ID NO 3; 29pp; English.

XX CC The invention relates to an isolated polynucleotide molecule encoding a
XX CC thermostable AvIII polypeptide. The polynucleotide is useful for
XX CC detection of a polynucleotide encoding AvIII. The polynucleotide is
XX CC useful for reducing cellulose in a starting material which involves
XX CC administering to the starting material, e.g. agricultural biomass or
XX CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
XX CC method further comprises administering a second polypeptide molecule
XX CC chosen from the glycoside hydrolase family of proteins. The present
XX CC sequence represents the amino acid sequence of Acidothermus
XX CC cellulolyticus avicelase AvIII catalytic domain.

XX SQ Sequence 740 AA;

Query Match 100.0%; Score 4036; DB 7; Length 740;
Best Local Similarity 100.0%; Pred. No. 8.4e-271;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATTQPYTWSNVAIGGGGFDGIVFNEGAPGILYVRDIDGMYRDAANGRWIPLLDWVGW 60
Db 1 ATTQPYTWSNVAIGGGGFDGIVFNEGAPGILYVRDIDGMYRDAANGRWIPLLDWVGW 60
Qy 61 NNWGYGVVSIADPINTNKWAAVGMYNDSMDPNDGAILRSSDQAGATWITPLPFKLG 120
Db 61 NNWGYGVVSIADPINTNKWAAVGMYNDSMDPNDGAILRSSDQAGATWITPLPFKLG 120
Qy 121 NMPGGMGERLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNPPDVGTYIANPDT 180
Db 121 NMPGGMGERLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNPPDVGTYIANPDT 180
Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNNPVFWRDGGATQAVGAPTGF 240
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNNPVFWRDGGATQAVGAPTGF 240
Qy 241 IPHKGVPDPVNHVLYIATSNTPGPDGSSGDKWKFVSTGVTWTRISVPBSTDTANDYFGY 300
Db 241 IPHKGVPDPVNHVLYIATSNTPGPDGSSGDKWKFVSTGVTWTRISVPBSTDTANDYFGY 300
Qy 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRYVLDISAE 360
Db 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRYVLDISAE 360
Qy 361 PMLTFGVQPNPVPSPKLGMDDEAMADPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
Db 361 PMLTFGVQPNPVPSPKLGMDDEAMADPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
Qy 421 APWKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY 480
Db 421 APWKGLEETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY 480
Qy 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKMFQGSERPGVTTGGTVAASADGSRFV 540
Db 481 AELNPSIIIVRAGSFDPSQPNDRHVAFTDGGKMFQGSERPGVTTGGTVAASADGSRFV 540
Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDDGGVTF 600
Db 541 WAPGDPQPVVYAVFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDDGGVTF 600
Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
Db 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
Qy 661 KSAPGSSYPVAVVGTGGVTVGAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRRV 720
Db 661 KSAPGSSYPVAVVGTGGVTVGAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRRV 720
Qy 721 YIGTNGRGIVYGDIGGAPSG 740
Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 4

ABP73015 ID ABP73015 standard; protein; 957 AA.

XX AC ABP73015;

XX DT 03-JUN-2003 (first entry)

XX DE Amino acid sequence of the avicelase AvIII.

XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
XX KW detergent; pulp processing; paper processing; feed processing; textile;
XX KW cellulose.

XX OS Acidothermus cellulolyticus.

XX	Key	Location/Qualifiers	
FH	Misc-difference 957		
FT	/note= "unspecified residue encoded by N"		
XX			
XX	W02003012090-A2.		
XX	13-FEB-2003.		
XX	28-JUL-2001; 2001WO-US023818.		
XX	28-JUL-2001; 2001WO-US023818.		
XX	(MIDE) MIDWEST RES INST.		
XX	Ding S, Adney WS, Vinzant TB, Himmel ME;		
XX	WPI; 2003-248177/24.		
DR	N-PSDB; ABZ77632.		
XX	New thermostable AviiiI peptide from Acidothermus cellulolyticus, useful for degradation of cellulose or in generating anti-AviiiI antibodies for purifying recombinant AviiiI polypeptides from genetically engineered host cells.		
XX	Claim 2; Page 20; 44pp; English.		
XX	The present sequence represents a thermostable avicelase polypeptide, designated AviiiI. AviiiI is a member of the glycoside hydrolase family of enzymes, and is a cellulase. AviiiI is useful in the conversion of biomass to biofuels and biofuel additives. It may be useful in the production of detergents, pulp and paper processing, food and feed processing and in textile processes. The thermostable AviiiI peptide is useful in the degradation of cellulose, and in generating specific anti-AviiiI antibodies that are useful in purifying recombinant AviiiI polypeptides from genetically engineered host cells, in detecting AviiiI polypeptide expression, as well as a reagent tool for characterizing the molecular actions of the polypeptide. The AviiiI polynucleotide is useful as a source of probes or primers in various diagnostic assays		
XX	Sequence 957 AA;		
XX	Query Match	100.0%; Score 4036; DB 6; Length 957;	
XX	Best Local Similarity	100.0%; Pred. No. 1.2e-270;	
XX	Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	ATTQPYTWSNVAIGGGGFDGIVFNEGAPGILYVRTDIGNMYRWDAAANGRWIPLLDWVGW 60	
Db	47	ATTQPYTWSNVAIGGGGFDGIVFNEGAPGILYVRTDIGNMYRWDAAANGRWIPLLDWVGW 106	
QY	61	NNWGYGVVSTAAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFKLG 120	
Db	107	NNWGYGVVSTAAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFKLG 166	
QY	121	NMPCRGGERLAVDPNNDNIIYFCAPSGKGLWRSTDGATWSQMTNPDVGTIYANPTDT 180	
Db	167	NMPCRGGERLAVDPNNDNIIYFCAPSGKGLWRSTDGATWSQMTNPDVGTIYANPTDT 226	
QY	181	TGYQSDIQGVVWVAFDKSSSLGQAKTIFVGVADPNPNVFWSRDGGATWQAVFGAPTGF 240	
Db	227	TGYQSDIQGVVWVAFDKSSSLGQAKTIFVGVADPNPNVFWSRDGGATWQAVFGAPTGF 286	
QY	241	IPHKGVDPVNVHLYIATSNTPGPGYDSSGDVWKFVSVTSGTWTTRISVPSTDTANDYFGY 300	
Db	287	IPHKGVDPVNVHLYIATSNTPGPGYDSSGDVWKFVSVTSGTWTTRISVPSTDTANDYFGY 346	
QY	301	SGLTIDRQHNTIMVATQISWMPDTIIFRSYDGGATWTRIDWTSYNRSLRYVLDISAE 360	
Db	347	SGLTIDRQHNTIMVATQISWMPDTIIFRSYDGGATWTRIDWTSYNRSLRYVLDISAE 406	
QY	361	PWLTFGVQPNPVPSPKLGWDEAMADPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420	
Db	407	PWLTFGVQPNPVPSPKLGWDEAMADPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466	
QY	421	APMVKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTFTSPVFTTGSVDY 480	
Db	467	APMVKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTFTSPVFTTGSVDY 526	
QY	481	AELNPSIIVRAGSFDPSQPNDRHVAFTDGGKWNFQSGEPGGVTTGGTVAASADGSRFV 540	
Db	527	AELNPSIIVRAGSFDPSQPNDRHVAFTDGGKWNFQSGEPGGVTTGGTVAASADGSRFV 586	
QY	541	WAPGDPGPVYVAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDCGVTF 600	
Db	587	WAPGDPGPVYVAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDCGVTF 646	
QY	601	QPVAAGLPSSGAVGMFHAVPKEGDMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660	
Db	647	QPVAAGLPSSGAVGMFHAVPKEGDMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706	
QY	661	KSAPGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720	
Db	707	KSAPGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 766	
QY	721	YIGTNGRGIVYGDIGGAPSG 740	
Db	767	YIGTNGRGIVYGDIGGAPSG 786	
XX	RESULT 5		
XX	ADD22921		
ID	ADD22921 standard; protein; 957 AA.		
XX	AC	ADD22921;	
XX	AC	ADD22921;	
DT	15-JAN-2004 (first entry)		
XX	Acidothermus cellulolyticus avicelase AviiiI.		
DE	enzyme; AviiiI; cellulose reduction; agricultural biomass;		
XX	municipal solid waste; glycoside hydrolase; avicelase.		
KW	Acidothermus cellulolyticus.		
XX	Key	Location/Qualifiers	
FH	Misc-difference 957		
FT	/label= Unknown		
FT	/note= "Encoded by N"		
XX	US2003108988-A1.		
XX	12-JUN-2003.		
XX	18-OCT-2002; 2002US-00155400.		
XX	28-JUL-2001; 2001US-00917376.		
XX	(DING/) DING S.		
PA	(ADNE/) ADNEY W S.		
PA	(VINZ/) VINZANT T B.		
PA	(HIMM/) HIMMEL M E.		
XX	Ding S, Adney WS, Vinzant TB, Himmel ME;		
XX	WPI; 2003-810853/76.		
DR	N-PSDB; ADD22922.		
XX	New isolated thermal tolerant avicelase polynucleotide useful for detection of a polynucleotide encoding AviiiI and for reducing cellulose in a starting material, e.g. municipal solid waste.		
XX	Claim 16; SEQ ID NO 1; 29pp; English.		
XX	The invention relates to an isolated polynucleotide molecule encoding a thermostable AviiiI polypeptide. The polynucleotide is useful for detection of a polynucleotide encoding AviiiI. The polynucleotide is		

CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents the amino acid sequence of Acidothermus
CC cellulolyticus avicelase Aviii.

XX SQ Sequence 957 AA;
Query Match 100.0%; Score 4036; DB 7; Length 957;
Best Local Similarity 100.0%; Pred. No. 1.2e-270;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
DB 47 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106
QY 61 NNMGYNGVVSIAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLKGG 120
DB 107 NNMGYNGVVSIAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLKGG 166
QY 121 NMPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDT 180
DB 167 NMPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDT 226
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 240
DB 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 286
QY 241 IPHKGVFDPNVHLYIATNTGGPDGSGDVWKSQVTSQWTRISPVSTDTANDYFY 300
DB 287 IPHKGVFDPNVHLYIATNTGGPDGSGDVWKSQVTSQWTRISPVSTDTANDYFY 346
QY 301 SGLTIDRQHPNTIMVATQISWMPDITIFRSTGGATWTRIMDWTSPNRSRLRYVLDISAE 360
DB 347 SGLTIDRQHPNTIMVATQISWMPDITIFRSTGGATWTRIMDWTSPNRSRLRYVLDISAE 406
QY 361 PWLTFGVQNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTWKDSSGQIHI 420
DB 407 PWLTFGVQNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTWKDSSGQIHI 466
QY 421 APMVKGLETA VNDLISPPSGAPLISALGDLGFTHADVAVPSTTIFTPSVFTTGSVDY 480
DB 467 APMVKGLETA VNDLISPPSGAPLISALGDLGFTHADVAVPSTTIFTPSVFTTGSVDY 526
QY 481 AELNPSIIVRAGSFDPSQPNDRHVAFSTDGKKNWFGQSEPGGVTTGGTVAASADGRFV 540
DB 527 AELNPSIIVRAGSFDPSQPNDRHVAFSTDGKKNWFGQSEPGGVTTGGTVAASADGRFV 586
QY 541 WAPGDPGPVAVVAVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALNGNTPFYRSTDCGVTF 600
DB 587 WAPGDPGPVAVVAVFGNSWAAASQGVPAQAQIRSDRVNPKTFYALNGNTPFYRSTDCGVTF 646
QY 601 QPVAAGLPSSGAVGVNMFHVPKGEGLWLAASGLYHSTNGSSWSAITGVSAVNVGFG 660
DB 647 QPVAAGLPSSGAVGVNMFHVPKGEGLWLAASGLYHSTNGSSWSAITGVSAVNVGFG 706
QY 661 KSAPGSSYPAVFGVGTGGTGYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720
DB 707 KSAPGSSYPAVFGVGTGGTGYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 766
QY 721 YIGTNGRGIVYGDIGGAPSG 740
DB 767 YIGTNGRGIVYGDIGGAPSG 786

RESULT 6
ADO52313
ID ADO52313 standard; protein; 957 AA.
XX AC
XX ADO52313;
XX

DT 15-JUL-2004 (first entry)
XX Acidothermus cellulolyticus avicelase III (Aviii).
DE Thermostable cellulase; avicelase III; Aviii; cellulose reduction;
XX biomass degradation; ethanol formation; industrial chemical;
KW fabric treatment; enzyme.
KW Acidothermus cellulolyticus.
OS
XX
FH Key Location/Qualifiers
FT Peptide 1..34
FT /label= Signal_peptide
FT Protein 35..957
FT /note= "Acidothermus cellulolyticus mature avicelase III"
FT Domain 35..187
FT /note = Carbohydrate binding domain type III
FT Domain 231..870
FT /note = GH74 catalytic domain
FT Misc-difference 957
FT /label= Unknown
XX
PN US2004038334-A1.
XX
XX 26-FEB-2004.
XX
XX 28-JUL-2001; 2001US-00917376.
XX
XX 28-JUL-2001; 2001US-00917376.
XX
XX (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
XX (HIMM/) HIMMEL M E.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
XX WPI; 2004-203224/19.
DR N-PSDB; ADO52314.
XX
XX Novel thermostable Aviii polypeptide of glycoside hydrolase family and
PT isolated from Acidothermus cellulolyticus, useful for degrading treated
PT biomass into simpler forms of carbohydrate.
XX
XX Claim 12; SEQ ID NO 1; 19pp; English.
XX
XX The invention relates to a thermostable cellulase enzyme, avicelase III
CC (Aviii) and its nucleic acid sequence. Aviii is useful for reducing
CC cellulose in a starting material. A thermostable Aviii peptide is useful
CC for degrading treated biomass into simpler forms of carbohydrate, which
CC is used in the formation of ethanol or other industrial chemicals. It is
CC also useful for treating fabrics to remove cellulose-containing stains.
CC The present sequence is Acidothermus cellulolyticus Aviii protein.

XX SQ Sequence 957 AA;

Query Match 100.0%; Score 4036; DB 8; Length 957;
Best Local Similarity 100.0%; Pred. No. 1.2e-270;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
DB 47 ATTQPTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106
QY 61 NNMGYNGVVSIAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLKGG 120
DB 107 NNMGYNGVVSIAADPINTNKVAAVGMVYNSWDNDGAILRSSDQATWQITPLPFLKGG 166
QY 121 NMPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDT 180
DB 167 NMPGRGMRGLAVDPNNNDILYFGAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDT 226
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNNPVFWSRDGGATWQAVPGAPTGF 240

Db 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPVFWSRDGGATWQAVPGAPTGF 286
QY 241 IPHKGVPDPVNHVLIATISNTGGPVDSGSDVWKEFSVTSGTWTTRISPPVSTDTANDYGY 300
Db 287 IPHKGVPDPVNHVLIATISNTGGPVDSGSDVWKEFSVTSGTWTTRISPPVSTDTANDYGY 346
QY 301 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGGATWTRIMDWTSPNRSRLRYVLDISAE 360
Db 347 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGGATWTRIMDWTSPNRSRLRYVLDISAE 406
QY 361 PMLTFGVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420
Db 407 PMLTFGVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 466
QY 421 APWKGLBEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIIFTSPTFTGTSVDY 480
Db 467 APWKGLBEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIIFTSPTFTGTSVDY 526
QY 481 AELNPSIIVRAGSFPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 540
Db 527 AELNPSIIVRAGSFPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 586
QY 541 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIQRSDRVNPKTFYALSNGTFYRSTDDGVTF 600
Db 587 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIQRSDRVNPKTFYALSNGTFYRSTDDGVTF 646
QY 601 QPVAAGLPSSGAVGVMFHAPVKGEDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
Db 647 QPVAAGLPSSGAVGVMFHAPVKGEDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706
QY 661 KSAPGSSYPVAVFVGTIGVTCAYSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720
Db 707 KSAPGSSYPVAVFVGTIGVTCAYSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 766
QY 721 YIGTNGRGIVYGDIGGAPSG 740
Db 767 YIGTNGRGIVYGDIGGAPSG 786

RESULT 7

AD052315

ID AD052315 standard; protein; 733 AA.

XX AD052315;

AC AD052315;

DT 15-JUL-2004 (first entry)

XX A. cellulolyticus avicelase III (AviIII) GH4 catalytic domain (CD).

XX Thermostable cellulase; avicelase III; AviIII; cellulose reduction;

KW biomass degradation; ethanol formation; industrial chemical;

KW fabric treatment; catalytic domain; CD.

XX Acidothermus cellulolyticus.

XX US2004038334-A1.

PN 26-FEB-2004.

XX 28-JUL-2001; 2001US-00917376.

XX 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

XX (ADNE/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

PI WPI; 2004-203224/19.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT Novel thermostable AviIII polypeptide of glycoside hydrolase family and
PT isolated from Acidothermus cellulolyticus, useful for degrading treated
PT biomass into simpler forms of carbohydrate.
XX Claim 28; SEQ ID NO 3; 19pp; English.

CC The invention relates to a thermostable cellulase enzyme, avicelase III
(AviIII) and its nucleic acid sequence. AviIII is useful for reducing
CC cellulose in a starting material. A thermostable AviIII peptide is useful
CC for degrading treated biomass into simpler forms of carbohydrate, which
CC is used in the formation of ethanol or other industrial chemicals. It is
CC also useful for treating fabrics to remove cellulose-containing stains.
CC The present sequence is Acidothermus cellulolyticus AviIII protein GH4
CC catalytic domain (CD).

SQ Sequence 733 AA;

Query Match 97.1%; Score 3920.5; DB 8; Length 733;
Best Local Similarity 99.1%; Pred. No. 8.4e-263;
Matches 733; Conservative 0; Mismatches 0; Indels 7; Gaps 7;

QY 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
Db 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
QY 61 NNWGVNGVVSIAADPINTNKVWAAVGMVYNSMDPNDGAILRSDGQATWQITPLPKLGG 120
Db 61 NNWGVNGVVSIAADPINTNKVWAAVGMVYNSMDPNDGAILRSDGQATWQITPLPKLGG 120
QY 121 NMPGRGMGERLAVDPNNDNIIYFGAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDT 180
Db 121 NMPGRGMGERLAV-PNND-ILYFGAPSGKGLWRSTDGATWQMTNFPDVGTYIANPTDT 178
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPVFWSRDGGATWQAVPGAPTGF 240
Db 179 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPVFWSRDGGATWQAV-PAPTGF 237
QY 241 IPHKGVPDPVNHVLIATISNTGGPVDSGSDVWKEFSVTSGTWTTRISPPVSTDTANDYGY 300
Db 238 IPHKG-FDPVNHVLIATISNTGGPVDSGSDVWKEFSVTSGTWTTRISPV-STDTANDYGY 295
QY 301 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGGATWTRIMDWTSPNRSRLRYVLDISAE 360
Db 296 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGGATWTRIMDWTSPNRSRLRYVLDISAE 355
QY 361 PMLTFGVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420
Db 356 PMLTFGVQPNPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 415
QY 421 APWKGLBEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIIFTSPTFTGTSVDY 480
Db 416 APWKGLBEETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTIIFTSPTFTGTSVDY 474
QY 481 AELNPSIIVRAGSFPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 540
Db 475 AELNPSIIVRAGSFPDPSSQPNDRHVAFSTDDGKNWFOGSEPGVTTGGTVAASADGSRFV 534
QY 541 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIQRSDRVNPKTFYALSNGTFYRSTDDGVTF 600
Db 535 WAPGDPGPVVYAVGFGNSWAASQGVPANAOIQRSDRVNPKTFYALSNGTFYRSTDDGVTF 594
QY 601 QPVAAGLPSSGAVGVMFHAPVKGEDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
Db 595 QPVAAGLPSSGAVGVMFHAPVKGEDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 654
QY 661 KSAPGSSYPVAVFVGTIGVTCAYSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720
Db 655 KSAPGSSYPVAVFVGTIGVTCAYSDDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 714
QY 721 YIGTNGRGIVYGDIGGAPSG 740
Db 715 YIGTNGRGIVYGDIGGAPSG 733

RESULT 8
 ABB99489
 ID ABB99489 standard; protein; 940 AA.
 AC
 XX
 AC ABB99489;
 XX
 DT 03-MAR-2003 (first entry)
 XX
 DE Amino acid sequence of a xyloglucanase enzyme.
 XX
 KW Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;
 XX textile scouring; cellulose fiber processing; rattling.
 OS Jonesia sp.
 XX
 PN WO200277242-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-DK000210.
 XX
 PR 27-MAR-2001; 2001DK-00000504.
 XX
 PA (NOVO) NOVOZMES AS.
 XX
 PI Duffner F, Sjolholm C;
 XX
 DR WPI; 2003-092855/08.
 XX
 DR N-PSDB; ABV76941.
 XX
 PT New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,
 XX and endogenous to a bacterium, useful in the textile industry for
 XX improving properties of cellulosic fibers, yarn or fabric.
 PS Claim 5; Page 72-75; 76pp; English.
 XX
 CC The present sequence represents a xyloglucanase enzyme, belonging to
 CC family 74 of glycosyl hydrolases. The enzyme is isolated from Jonesia sp.
 CC DSM14140. The enzyme is useful in processes for machine treatment of
 CC fabrics. It is also useful in the textile industry for improving the
 CC properties of cellulosic fibers, yarn, woven or non-woven fabric, and in
 CC a textile scouring process step. The xyloglucanase enzyme is also useful
 CC in the cellulose fiber processing industry for rating of fibers such as
 CC hemp, jute, flax and linen. It is useful for preventing binding of
 CC certain soils to the xyloglucan left on the cellulosic material
 CC
 SQ Sequence 940 AA;
 Query Match 51.1%; Score 2063; DB 6; Length 940;
 Best Local Similarity 52.1%; Pred. No. 5.9e-134;
 Matches 386; Conservative 112; Mismatches 225; Indels 18; Gaps 9;
 QY 6 YTSNVAIGGGFVGGFVGNVNEGAPILYVTRDIGMYRDAANGRWIPLLDVWGNWY 65
 DB 44 YSWSNVEIVGGFVPGIVFNQKDPGLVYARTDIGAYRLNDSTGRWIPLTDRHGDW 103
 QY 66 NGVVSIAADPINTNKVAAVGMVMTNSWDPNDCAILRSSDQATWITPLPFLGGMPCR 125
 DB 104 SGILSLATDPVDTNRYLAAGTYSDMDPQNGAILRSADKGETWKTMLPFRVGNMPCR 163
 QY 126 GNGERLAVDPNNNLIYFCAPSGKGLWRSTSDGATWSQMTNFPDVGTYIANPTDTTG-YQ 184
 DB 164 GNGERLAVDPNNNNVLYFCAESGNGLWSTDYGTWGVTSFPNAGNYVA---DASGAYT 220
 QY 185 SDIQGVWVAFPKSSSSQAKTIFVGVADPNNPVFWRDGGATWQVPGAPTGPPIPK 244
 DB 221 GQNGVWVWTFDPTSAKAGQTTQTIYGVADKQNNVYRSTDGGATWQRVPGQGTGFLAQK 280
 QY 245 GVFDPNVHLYIATNSVGPYSGSDVWKFVTSCTWTRISPVSTDTANDYFGYSLT 304
 DB 281 GVLDHKGOQLYIATSDTGPDGSDVWRLDISGQWTRISPIST-SSNSAFGYSGLA 339

QY 305 IDRHQNTIMVATQISWHPDTIIFRSTDCGATWTRIDWTSYSPNRSRLRYVLDISASWLT 364
 DB 340 IDRKNPDTIMVVSQVSWPDMYVYRSTDRKTSPIWELNGSQPRTKQYNHDSYGAPWLD 399
 QY 365 FGVPNPVPSPKLGWMDMAIDPNSDRMLYGTGATLYATNDLTQWDSGGQIHAPMV 424
 DB 400 FGNTAKEPEANPKLGWMTQSFIDPHNSDRFFYGAGIYGGTNLTNWDKXKVDITVKA 459
 QY 425 KGLETAVNDLISPPSGAPLISALGDLGFTHADVTAVPST-IFTSPVFTTGTSTVDYAE 483
 DB 460 QGIEETAQDLAAPPNGICLYSALADIGGFTKDISQVFNKYYKPHPHDVTISIDFAES 519
 QY 484 NPSIIVRAGSFDPSSQPNDRHVAFTDGGKNNPQSGSEPGVTTGGTVAASADSRFWAP 543
 DB 520 KPATVVRAGK--SIGETTSTWVGSTDAGETWKPATSGVKGPGSITVSANASSIWWAP 577
 QY 544 GDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKFTFVLSNGTFFYRSTDCGVTPOV 603
 DB 578 --EGAAPRRSTNSGSSWSTVSGLPNAQVASDRVNANTLYGFDGKFFYHSTNGGASFTAS 635
 QY 604 A-AGLPSSGAVGMFHAVPGKEGDLMLA-----ASSGLYHSTNGSSWSAITGVSSAVN 657
 DB 636 AFTGPTSG--NVRFRAVPRQGHLMWLAGVSGSTYGMWRSTDGKNTKVSQVQEGDAV 693
 QY 658 GFGKAPGSSYPVAVFVVGTTGGVYRSDDCGTTWVLINDQHQYGNWGOAITGDHANL 717
 DB 694 GFGKATSSGYPVIFTSAKIDGVRGIFRSDDCKTWKRINDNHNQWTTGASITGDPDVE 753
 QY 718 RRVYGTNGRGIVYGDIGAP 738
 DB 754 GRVYGTNGRGIVGDSSTPP 774
 RESULT 9
 ADD22927
 ID ADD22927 standard; protein; 726 AA.
 XX
 AC ADD22927;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Aspergillus aculeatus avicelase III catalytic domain.
 XX
 KW enzyme; AvIII; cellulose reduction; agricultural biomass;
 KW municipal solid waste; glycoside hydrolase; avicelase.
 XX
 OS Aspergillus aculeatus.
 XX
 PN US2003108988-A1.
 XX
 PD 12-JUN-2003.
 XX
 PF 18-OCT-2002; 2002US-00155400.
 XX
 PR 28-JUL-2001; 2001US-00917376.
 XX
 PA (DING/) DING S.
 PA (ADNE/) ADNEY W S.
 PA (VINZ/) VINZANT T B.
 PA (HIMM/) HIMMEL M E.
 XX
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;
 XX
 DR WPI; 2003-810853/76.
 XX
 PT New isolated thermal tolerant avicelase polynucleotide useful for
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose
 PT in a starting material, e.g. municipal solid waste.
 XX
 PS Example 2; SEQ ID NO 7; 29pp; English.
 XX
 CC The invention relates to an isolated polynucleotide molecule encoding a
 CC thermostable AvIII polypeptide. The polynucleotide is useful for


```
QY 300 YSLGTIDRQHPNTIMVATQISWMPDITIFRSTDCGATWTRIDWTSYPNRSLRYVLDISA 359
DB 314 FGLGLDLQKPTGLVASLNSWPDQAQLFRSTDSGTWSPINAWASYPTETYYISITPK 373
QY 360 EFWLTFG-VQNPFPVPS-----PKLGMDEAMADPFNSDRMLYGTGATLYATNDLTWKDS 414
DB 374 APWIKNFIDVTSSEPSDGLIKRLGWMTESLEIDPTDSNHMLYGTGWTIFGCHDLTNWDT 433
QY 415 GGOIHAPWVKGLEETAVNDLISPPSGAPLISALDGLGFTHA---DVTAVPSTIFTSPV 471
DB 434 RHNVSIOQLADGIEEFSVODLASAPGSELLAAVGDNDNGFTFASRNDLGTSPQTVWATPT 493
QY 472 FTTGTSVDYAEALNPSIIVRAGSFDPSSQPNDRHVAFTDGGKNWFGQSEPGVTTGGTVA 531
DB 494 WATSTVDYAGNSVKSVVRVNTAGTQO-----VAISDGGATWIDYAADTSMNGGIVA 548
QY 532 ASADGSRFVWAPGDPGPVYAVFGFNSWAASQGVPANAIQIRSDRNPKTFFYALNSGTFF 591
DB 549 YSADGDTILWSTASSG---VQRQFQGSFASVSLPAGAVIASDKKTNVSVFYAGSGSTFY 605
QY 592 RSTDDGVTFQPVAAGLPSSGAVGM--FHAVPGKEGDLWLAASSGLYHSTNGGSSWSAL- 648
DB 606 VSKDTGSSSF---TRG-PKLGSGAGTIRDAHPPTTAGTLYVSTDVGIFRSTDSGTTFGQVS 661
QY 649 TGVSSAVNVFGKSAFGSSYPVAVFVVGTTGGVTGA--YRSDCGTWTWVLINDDQHQYGNW 706
DB 662 TALNTYQIALGVGS-GSNW-NLYAFGT--GFSGARLYASGDSGASWTIDIQSQQFGSID 717
QY 707 GQAITGDHANLRVYVIGTNGRIVY--GDIGAPSG 740
DB 718 STKVAGSGTAGVYVINGRGGVFAQGVGGGTGG 753

RESULT 11
ADH51581
ID ADH51581 standard; protein; 818 AA.
AC ADH51581;
XX
XX 25-MAR-2004 (first entry)
XX
XX Trichoderma reesi EGVI protein sequence related to ethanol production.
KW EGVI protein; endoglucanase; ethanol production; biomass composition;
KW fermentation; cellulohydrolase; detergent production; softening agent;
KW cotton fabric; wood pulp degradation; sugar; enzyme.
XX
XX Hypocrea jecorina.
XX
XX Key Location/Qualifiers
XX Misc-difference 1..818
XX /label= OTHER
XX /note= "OTHER= All Xaa's given in this sequence are
XX unknown amino acids, given as J in the specification.
XX they should probably be Ile."
XX
XX US2003113732-A1.
XX
XX 19-JUN-2003.
XX
XX 18-DEC-2001; 2001US-00026994.
XX
XX 18-DEC-2001; 2001US-00026994.
XX
XX (DUNN/) DUNN-COLEMAN N.
XX (GOED/) GOEDEGEBUUR F.
XX (WARD/) WARD M.
XX (YAOJ/) YAO J.
XX
XX Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;
XX WPI; 2004-106460/11.
XX
XX N-PSDB; ADH51580, ADH51583.
XX
```

```
XX
PT New substantially purified EGVI polypeptide with the biological activity
PT of endoglucanase, useful for degrading biomass to ethanol.
XX
PS Claim 18; SEQ ID NO 2; 26pp; English.
XX
XX This invention relates to a novel purified EGVI protein with the
CC biological activity of an endoglucanase, comprising an amino acid
CC sequence chosen from a fully defined sequence as given in specification.
CC The protein is useful for producing ethanol, which involves contacting a
CC biomass composition with an enzymatic composition containing the protein
CC of the invention to yield a sugar solution, adding to the sugar solution
CC a fermentative microorganism, and culturing the fermentative
CC microorganism under conditions sufficient to produce ethanol, where the
CC biomass composition may be optionally pretreated. The method further
CC involves the addition of at least one endoglucanase or cellulohydrolase.
CC The pretreatment is with a dilute acid. The protein of the invention is
CC also useful for the production of a detergent composition which may be
CC used as softening agent and for improving the feel of cotton fabrics and
CC for degrading wood pulp into sugars. The present sequence is that of the
CC T reesei EGVI protein of the invention.
XX
XX Sequence 818 AA;
XX
XX Query Match 38.6%; Score 1557.5; DB 8; Length 818;
XX Best Local Similarity 42.6%; Pred. No. 5.2e-99;
XX Matches 322; Conservative 120; Mismatches 269; Indels 45; Gaps 20;
XX
QY 6 YTSNVAI-GGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRIPLLDWV---GW 60
DB 2 FSKNVKLGCGGFGVPIIHPKTKGVAVARTDIGLYRLN-ADDSWATVTDGIADNAGW 60
QY 61 NNWYNGVVSIAADPINTKNVAAVGMVYNSWPNDCGAILRSSDOGATWQITPLPFLKGG 120
DB 61 HNW---GIDAVALDPQDDQKVAAVGMVYNSWPNDCGAILRSSDOGATWQITPLPFLKGG 117
QY 121 NMPGRGMGERLAVDPNNNDILYFGAPSGKGLWSTDSGATWSQNTNPPDVGTYIANPTDT 180
DB 118 NMPGRGAGERLAVDPANSNIYFGRSGNGLWSTDSGATWSQNTNPPDVGTYIANPTDT 177
QY 181 TGYQSDIQGVVWVAFDKSSSLQOAKTIFGVVADP--NNPVFWSRSDGATWQAVPGAPTQ 239
DB 178 NGVNSDKQGLMWVTFDSTSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPK 237
QY 240 FPHKGVDVFNHVLIAITNSCTGPDGSSGVWKFVSTGWTTRISPPSTDTANDYFG 299
DB 238 YFPFKAKIQPAEKALYLTYSDDGTGPDYDGLGTVWVRDIAGGTWDXDTPVSGSDL---YFG 294
QY 300 YSGLTIDRQHPNTIMVATQISWMPDITIFRSTDCGATWTRIDWTSYPNRSLRYVLDISA 359
DB 295 FGLGLDLQKPTGLVASLNSWPDQAQLFRSTDSGTWSPINAWASYPTETYYISITPK 354
QY 360 EFWLTFG-VQNPFPVPS-----PKLGMDEAMADPFNSDRMLYGTGATLYATNDLTWKDS 414
DB 355 APWIKNFIDVTSSEPSDGLIKRLGWMTESLEIDPTDSNHMLYGTGWTIFGCHDLTNWDT 414
QY 415 GGOIHAPWVKGLEETAVNDLISPPSGAPLISALDGLGFTHA---DVTAVPSTIFTSPV 471
DB 415 RHNVSIOQLADGIEEFSVODLASAPGSELLAAVGDNDNGFTFASRNDLGTSPQTVWATPT 474
QY 472 FTTGTSVDYAEALNPSIIVRAGSFDPSSQPNDRHVAFTDGGKNWFGQSEPGVTTGGTVA 531
DB 475 WATSTVDYAGNSVKSVVRVNTAGTQO-----VAISDGGATWIDYAADTSMNGGIVA 528
QY 532 ASADGSRFVWAPGDPGPVYAVFGFNSWAASQGVPANAIQIRSDRNPKTFFYALNSGTFF 591
DB 529 YSADGDTILWSTASSG---VQRQFQGSFASVSLPAGAVIASDKKTNVSVFYAGSGSTFY 585
QY 592 RSTDDGVTFQPVAAGLPSSGAVGM--FHAVPGKEGDLWLAASSGLYHSTNGGSSWSAL- 648
DB 586 VSKDTGSSSF---TRG-PKLGSGAGTIPDIAAHPPTTAGTLYVSTDVGIFRSTDSGTTFGQVS 641
QY 649 TGVSSAVNVFGKSAFGSSYPVAVFVVGTTGGVTGA--YRSDCGTWTWVLINDDQHQYGNW 706
XX
```


Db	642	TALNTVQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASWTDIQSGQSGSID	697	Db	173	S-----VIFDP-----ERNGTIYASATAPQG-MYTHDGGVSWEPVAGQSPSWLNR	218
Qy	707	GOAITGDHANLRRVYICNGRGIYV--GDIGAPSG	740	Qy	242	-----PHKGVDPVNVHLYIATSNLTGPGYDSSGDMWKFVSFTSGTWTRI-	285
Db	698	STKVAGSGTAGQVYVGTNGRGVFAQGTGGTGG	733	Db	219	TGAPDKKPSIAPOPMKVALTP--NPLYVYADYPGWGVTFGEVWRQNRISGAWDDIT	276
RESULT 12				Qy	286	-----SPVSTDTANDYFGYSGLTIDROHPNTIMVATOISWMDPTI---IFRSTDTGGAT	336
ADH19120				Db	277	PRVGNSSPAPYNNQTFPAGGFCGLSVDATNPRLVIT-LDRDPGAPALDSIYLSLTDAGAT	335
XX	ID	ADH19120 standard; protein; 789 AA.		Qy	337	WTRIDWTS-----YPNRSURYVLDISAEPMWTFQVQP-----NPPVPSP---	379
XX	AC	ADH19120;		Db	336	WKDVTQLSSPSNLEGNWGHPTNAARY-KDGPVPWLDFNNGPQWQGYGAPHGTPGLTKFG	394
XX	DT	11-MAR-2004 (first entry)		Qy	380	WMDEAMAIIDPNSDRMLYGTGATLYATNDLTK-----WDSGGQIHIAP-----MVKGLEETA	431
XX	DE	Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme mat protein.		Db	395	WWMGAVLIDPFPNPEHMYGTGATIWATDLSRVEKDW-----APSWYLQIDGIEENA	446
XX	KW	xyloglucan oligosaccharide degradation; enzyme; genetic engineering;		Qy	432	VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDVYAEINPSIIVRA	491
XX	KW	mature.		Db	447	ILSLRSPKSGAALLSGIGDISGMKDDLT-K-PQKMGAPQFNSLDSIDAAGNFPNVVRA	505
XX	OS	Geotrichum sp. M128.		Qy	492	GSPDPSSQPNDRHVAFTDGGKNW--FQSGEPGGVTT---GGTVAASADGSRFVWAP--G	544
XX	FN	EP1350844-A2.		Db	506	GSSGHEYDSACARGAYATDGGDAWTFPTCPPEGMNASHYQGSTIAVDASGSQIVWSTKLD	565
XX	PD	08-OCT-2003.		Qy	545	DPGQPVVYAVGFGNSWAASQGVPA-----NAQIRSDRVNPKTFVALSNGTFRVSTDDGV	598
XX	PF	25-MAR-2003; 2003EP-00251866.		Db	566	EQASGPWYSHDYGKTWS-----VPAGDLKAQTANVLSKVDQDTFYATDGGKFFVSTDDGK	621
XX	PR	25-MAR-2002; 2002JJP-00083433.		Qy	599	TFQPVAAGLPSSGAVGVNMFHAPVPGKEGDLWL-AASSGLYHSTNGSGSSSAI-TGVSSAYN	656
XX	PA	(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.		Db	622	SYAAKAGLVN--GTSMLPAVNPVWAGDVWVPVPSGLFHSSTDGASFTRVGTANATLVS	679
XX	PI	Yaoi K, Mitsuishi Y;		Qy	657	VGFGKS-----APGSSYPVAVVVGTT--IGSVTGAYRSDDCGTTWVLINDDQHOYGNWGOAI	710
XX	DR	WPI; 2004-100948/11.		Db	680	VGAPKSKDGKASAPSAVFIWGTDPKGSIDIGLYRSDDNGSTWTRVNDQEHYSG-PTMI	738
XX	DR	N-PSDB; ADH19119.		Qy	711	TGDHANLRVYIGTNGRGIVYDI	734
XX	PT	Novel xyloglucan oligosaccharide-degrading enzyme with a different		Db	739	EADPKVYGRVYLTNGRGIVYADL	762
XX	PS	degradation mechanism from known enzymes.					
XX	PS	Claim 2; SEQ ID NO 14; 39pp; English.					
XX	CC	The invention relates to a novel xyloglucan oligosaccharide-degrading					
XX	CC	enzyme with a different degradation mechanism from known enzymes. The					
XX	CC	primary structure and polynucleotide structure of the xyloglucan					
XX	CC	oligosaccharide-degrading enzyme provided by the present invention allows					
XX	CC	high-purity polypeptides having a xyloglucan oligosaccharide-degradation					
XX	CC	activity to be prepared at a low cost through a genetic engineering					
XX	CC	process. The current sequence is that of the Geotrichum sp. M128					
XX	CC	xyloglucan oligosaccharide-degrading enzyme mature protein of the					
XX	CC	invention.					
XX	SQ	Sequence 789 AA;					
Query Match		27.7%; Score 1117; DB 8; Length 789;					
Best Local Similarity		35.0%; Pred. No. 1.7e-68;					
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;							
Qy	6	YTNWVAIGGGGVDFDGIYFNEGAPILYVRDIDGMYRWDAAANGRWIPLLDWGVNNGWY	65				
Db	4	YEFKVAIGGGGTYTGIVAHPTKOLLYARDIDGAYRWADAGTSKWIPNDIFIAQDNWI	63				
Qy	66	NGVYSIADPINTNKVMAAVGMYT-NSWDPNDGAILRSSDQAGATWQITPFLPKLGGNMPG	124				
Db	64	MGTESIALDPNPNDRLYLAQRYGVDEW---AAPVYVEDRGSQSFYIESPPFGANDMG	119				
Qy	125	RGMERLAVDPNNNINILYFGAPSGKGLWRSTDSGATWSQMTNPPDPVGYIANPTDTTGYQ	184				
Db	120	RNNGERLAVNPNFNSNEVWMTGT-EGIKWSDRAKTWTNVTISIPDAF-----TNGIGYT	172				
Qy	185	SDIOGVVWVAFDKSSSSISQASKTIFVGVGAPNPNFVWSRDGATWQAVPGAPTGF---	241				

RESULT 13
ADH19118
ID ADH19118 standard; protein; 812 AA.

AC ADH19118;

DT 11-MAR-2004 (first entry)

XX Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme pre protein.

XX xyloglucan oligosaccharide degradation; enzyme; genetic engineering;
KW precursor.

XX Geotrichum sp. M128.

XX EP1350844-A2.

XX 08-OCT-2003.

XX 25-MAR-2003; 2003EP-00251866.

XX 25-MAR-2002; 2002JJP-00083433.

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

XX Yaoi K, Mitsuishi Y;

XX WPI; 2004-100948/11.

XX DR N-PSDB; ADH19117.

QY 242 -----PHKGVDPVNVHLYIATSNITGGSPYDSSGDWKFVSITSGTWRI- 285
Db 220 TGAFFDKKPKASIAQPMKVALTP--NFLVYVADYDGPVGWTFKVMQRNRTSGAWDDIT 277
QY 286 -----SVPSTDTANDYFGSLGTLIDRQHPNTIMVATQISWPDPTI---IFRSTDGAT 336
Db 278 PRVGNSSPAPNNQTFPAGGFCGLSDATNPNRLVWIT-LDRDFGPALDSIYLSYLDAGAT 336
QY 337 WTRLDWTS-----YPNRSLRYLVDISAEPWLTGVP-----NPPVPSP---KL 379
Db 337 WKDVTQLSSPNLEGNWGHPTNAARY-KDGPVPLDFFNNGPQGGYCAPHGTFGLTKFG 395
QY 380 WMDAMAIIDPNSDRMLYGTGATLYATNDLTK---WDSGGQIHIAP-----MVKGLEETA 431
Db 396 WMSAVLIDPNEHLMYGTGATIWATDLSRVEKDW-----APSWLIQDIEBNA 447
QY 432 VNDLISPPSGAPLISALDGLGGFTHADVAVPSTIFTSPTFTGTSVDYAEINPSIIVRA 491
Db 448 ILSLRSPKSGAALLSGIGDISGMKHDDLT-K-PQRMFGAPQFSNLDISIDAAGNFPNVVRA 506
QY 492 GSFDPSSQPNDRHVAFSYDGGKNW--FQGSBPGGVTT---GGTVAASADGSRFWAP--G 544
Db 507 GSSGHEYDSACARGATADGGDAMTIFPTCPGMNASHYQGSTIAVDASGSGIWNSTKLD 566
QY 545 DPGQPVVYAVGFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFFYRSTDCGV 598
Db 567 EQASGPWYSHDYGKTS-----VPAGDLKAQATANVLSDKQDGTFFATDGGKFFVSTDGK 622
QY 599 TFQPVAAGLSPSGAVGVNMFHAPVQKGBDLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN 656
Db 623 SYAAKAGGLVT--GTSLMPAVNPNVAGDVVFPVEGGLFHSSTDFGASPTRVGTANATLVS 680
QY 657 VGFGKS-----APGSSYPVAVFVGT--IGGVTGAYRSDDCGTTWVLINDDOHOYGNWGOAI 710
Db 691 VGAPKSKSDGKKASAPSAVFVWGTDKPGSDIGLYRSDDNGSGTWRVNDQEHNSG-PTMI 739
QY 711 TGDHANLRRVYIGTNGRGIYVYDI 734
Db 740 EADPKVYGRVILGTNGRGIYVADL 763
RESULT 15
ADR90298
ID ADR90298 standard; protein; 756 AA.
XX AC ADR90298;
XX DT 16-DEC-2004 (first entry)
XX DE Geotrichum sp. M128 xyloglucanendohydrolase mature protein.
XX KW endo-type xyloglucanendohydrolase;
KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
KW plant cell differentiation; mature; enzyme.
XX OS Geotrichum sp. M128.
XX PN JP2004261037-A.
XX PD 24-SEP-2004.
XX PF 28-FEB-2003; 2003JP-00053286.
XX PR 28-FEB-2003; 2003JP-00053286.
XX PA (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.
XX DR WPI; 2004-665466/65.
XX DR N-PSDB; ADR90297.
PT Novel xyloglucan hydrolase derived from Geotrichum genus, having ability to hydrolyze beta-1, 4-glycoside linkage of glucose residue in

PT xyloglucan, useful for elucidating structure or function of xyloglucan.
XX Claim 2; SEQ ID NO 15; 86bp; Japanese.
XX CC The invention relates to a novel xyloglucanendohydrolase (xyloglucan-specific endo-b-1,4-glucanase) enzyme derived from the Geotrichum genus and having the ability to hydrolyse the beta-1,4-glycosidic linkage of a glucose residue, but not that of a xylose residue contained in a xyloglucan. The polypeptide of the invention may be useful for elucidating the structure or function of a xyloglucan having an important role in the differentiation of a plant cell. The current sequence is that of the Geotrichum sp. M128 xyloglucanendohydrolase mature protein of the invention.
XX Sequence 756 AA;
QY Query Match 24.3%; Score 979; DB 8; Length 756;
Best Local Similarity 33.1%; Pred. No. 5.7e-59;
Matches 264; Conservative 122; Mismatches 279; Indels 132; Gaps 33;
QY 11 VAIGGGFVDGIVFNEGAPGILYVTRDIGMYRMDAANGRWIPLLDWV-----GMNNWGY 65
Db 8 VTISGGFISGLVAHPTEKDLIYARTDIGTYRNAAKWEWEPITDFIINNALAGNANL 67
QY 66 NGWVSIADPTNTNKKVAAVGMVYNSWPDNGAILRSSDQGWATQITPLPFKLGMPGR 125
Db 68 LGTESIALDPNDRLYLAQGDYV-QMDP-WAAFLVSDDRGKTFKQYRSPVPMGANDMGR 125
QY 126 GMBERLAVDPNNDILYFAGSGKGLWSTDSGATWSQMTNPPDPVGTIANTTDTTGYOS 185
Db 126 NGERLAVNPHWTDELWFGSRT-QGLWRSTDRQAQWNRNQLPDSSTY-----GI-- 174
QY 186 DIQGVVWVAFPSGSSSLGQASKTIFVGVADPNPNPVFWSRDGGATWQAVPGAPT----- 238
Db 175 ---GLISVIFDPKNGVTAYASHAVGGLW-----VTW--DGGANWSQVGGQTQNSDWTK 224
QY 239 -----GFIPHKGVDPVNVHLYIATSNITGGSPYDSSGDWKFVSITSGTWRI 285
Db 225 SIVAASGTAIQSSGFLPIKIALGK-NGRLYITYSDAPGFWGVLYGEVMSYDPTNGNWKHI 283
QY 286 SPVSTDTANDY-----FQYSLGTLIDRQHPNTIMVATQISWPDPTIIFRSTDCG 334
Db 284 T--PSREGANTYPAPTGNKKVVPVGGWNGISVG--NGDTVVVSTLDANGSDS-VYLSRDAG 338
QY 335 ATWTRIDWTSYPNRSRLRYLVDISA-----PMLTFGVQPNPPVSPSKLGWMDAMAI 387
Db 339 NSWKDLGKLTTPAGAGNSQKESDAKLRNGTLPWLSFQNRGSGIV---GFGWLLAAILL 395
QY 388 DPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPM-----VKGLEETAVNDLISPPSG-A 442
Db 396 DFP-SDRLLYGTGAVIATDAVSRADS-----NQAFSWYINTEGIEETAILVLKSPGAPGA 450
QY 443 PLISALGDLGGFTHADVAVPSTIFTSPTFTGTSVDYAEINPSIIVRAGSFD--PSSQP 500
Db 451 HLFSGMYDLGMRHDDF-SVPQPMYSKPTFSSTDGLDFAGRAANLARYGRNDHPDAGVA 509
QY 501 NDRHVAFSTDGGKNW--FQGSBP-GVTTGTGTVAASADGSRFWMAPGD-PGQPVVYAVGP 556
Db 510 GCTQGYATTNGSDSWTLFTQTCVPSLEVNGNGTIVAGADKTFVWSFSAKDGKPYTSSDY 569
QY 557 GNSWAASQGVPANAO-IRSDRVNPKTFYALSNGTFFYRSTDCGVTTFQPVAAGLFPSSGAVG 615
Db 570 GKTWTAPSGLSKQTTGIAADRQANTFVYVYEGDFVSTDCGSKSYTKKGNGLPCCW--- 626
QY 616 MFHAVPG-----KEGDLWLAAS-GLYHSTNGSSWSAITGVSSAVNVPVGFSAKSGSSYPA 670
Db 627 -YTGTFVTSNLRAGELWVSKVGVIYHSTDFGNTFTALAGSSSLN-----PA 673
QY 671 VFWVGT-----IGGV-----TGAYRSDDCGTTWVLINDDOHOYGNWQAITGD 713
Db 674 VFSIGAPQTPNATETFLWGLIPSAQSQPEGLYMSDNGGLWTRLNDDAHNYGG-ATVISGD 732
QY 714 HANLRRVYIGTNGRGI 730

Db 733 PRIYGRVYIGNMGRGII 749
||||| |||||

Search completed: October 5, 2005, 08:03:10
Job time : 145.761 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 5, 2005, 07:48:13 ; Search time 20,5336 Seconds
(without alignments)
3467.504 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTQPYTWSNVAIGGGFVD.....YGTNGRGIVYDIGGAPSG 740

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2371	58.7	890	2 T35237	probable secreted
2	2009	49.8	839	2 D97013	probably secreted
3	1680	41.6	856	2 T00349	Avicelase III - As
4	1160.5	28.8	707	2 F72393	hypothetical prote
5	238	5.9	2468	2 A83412	hypothetical prote
6	202.5	5.0	2174	2 E95965	hypothetical glyci
7	189	4.7	3972	2 S75251	hypothetical prote
8	186.5	4.6	4199	2 S76412	hypothetical prote
9	182	4.5	1904	2 T13256	tail-host specific
10	175.5	4.3	2523	2 F70846	probable PPE prote
11	172	4.3	993	2 AE1905	outer membrane sec
12	171	4.2	908	2 AE2254	hypothetical prote
13	170.5	4.2	5188	2 B85547	probable RTX famil
14	168.5	4.2	5291	2 F90696	hypothetical prote
15	168	4.2	2124	2 A28452	proteoglycan core
16	167.5	4.2	2554	2 AB3528	extracellular seri
17	166.5	4.1	902	2 H87323	hypothetical prote
18	166.5	4.1	980	2 H90681	probable flagellin
19	166	4.1	699	2 D70533	hypothetical prote
20	166	4.1	1468	2 A44345	nucleoporin - rat
21	165.5	4.1	980	2 D85532	probable structura
22	163.5	4.1	618	2 T49741	related to stress
23	163.5	4.1	13055	2 T16580	hypothetical prote
24	163	4.0	1684	2 S10789	amyase A-180 - al
25	160.5	4.0	3716	2 E70969	probable PPE prote
26	160	4.0	1341	2 H98323	hypothetical prote
27	160	4.0	3624	2 AD8035	large repetitive p
28	158.5	3.9	1441	2 B86807	hypothetical prote
29	158	3.9	2232	2 T34434	hypothetical prote

30	157	3.9	809	2 A55547	quinate-shikimate
31	156.5	3.9	2204	2 A70524	probable PPE prote
32	156	3.9	3157	2 B70969	probable PPE prote
33	155	3.8	4180	2 G83559	hypothetical prote
34	154.5	3.8	1032	2 T34433	hypothetical prote
35	154.5	3.8	3570	2 T45025	mucin MUC5B, trach
36	153	3.8	1055	2 A87364	OmpA-related prote
37	153	3.8	1821	2 AG2335	hypothetical prote
38	153	3.8	13288	2 T30309	mucin, submaxillar
39	152.5	3.8	3016	2 S77300	hypothetical prote
40	152	3.8	348	2 T35248	probable oxidoredu
41	152	3.8	1049	2 T42045	beta transducin-li
42	152	3.8	1196	2 A29130	beta-amylase (EC 3
43	152	3.8	3472	2 T31308	hypothetical 367K
44	151.5	3.8	820	2 B72575	hypothetical prote
45	151.5	3.8	1145	2 B75625	hypothetical prote

ALIGNMENTS

RESULT 1

T35237
Probable secreted cellulase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: T35237
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21572
A:Accession: T35237
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-890 <SEE>
A:Cross-references: UNIPROT:O86727; EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GN00070; SCOE
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOE:SC5C7.30C

Query Match 58.7%; Score 2371; DB 2; Length 890;
Best Local Similarity 57.7%; Pred. No. 5.9e-131;
Matches 431; Conservative 109; Mismatches 183; Indels 24; Gaps 10;

QY	6	YTSNVAIGGGFVDGIVFNEGAPGILYVRTDIDGMYRWDANGRWITPLLDVGMNNGY	65
DB	41	YTKNARIIDGGFVPGIVFNRTEKDLAYARTDIGGAYRWOEEHWTPLLDHVGWDWGH	100
QY	66	NGVVSIAADPINTNKYMAAVGMYTNSWDPNDDGAILRSSDQATWQITPLPFLGGMNPGR	125
DB	101	TGVVALASDAVDPRVYAAVGTITNDWDPNGAVLRSADRGSWEKADLPFLGGMNPGR	160
QY	126	GMGERLAVDPNNNIIYFGAPSGKGLWRSTDGATWSQMTNFFDVGVYIANPTDTTGVQS	185
DB	161	GMGERLAVDPHDNDVLYLGAPSGHGLWRSTDAGVTWSEVTAFPNPGNYAQDPNDTSGVAS	220
QY	186	DIQGVVWVAPDKSS-SSLGQASKTIIFGVADPNPVPWSDGGATWQAVPCAPGFIPIHK	244
DB	221	DNQGITVWTPDESTGGGAGTATRTLYYGVADKENAVYRSTDAGATWERLAGOQTYLAHK	280
QY	245	GVFDPVNHVLYIATNSGTGSGDGVWKFVSITSGTWTTRISVPVSTDTANDYFGYSGLT	304
DB	281	GVLDAGNYLYLAYSDTGGYDGGKLYIATATGTWTDISPAAEADT---YGFSGLT	337
QY	305	IDRQHPNTIMVATQISWPDITIIFRSTDGGATWTRIDWTSYPNRSRYVLDISAEPMLT	364
DB	338	VDRQRPCTVWATYSSWWPDITQIFRSTDGATWSQAWSYTSYPDRENRKTYMDVSSPMLT	397
QY	365	FGVQPNPPVPSPKLGWMDAEMADIPFNSDMLYGTGATLYATNDLTW-DSSGGQIHIAPM	423
DB	398	WGANPAPPEQTPKLGWMTALEIDPFDSDRMVYGTGATVGTENTLTWMDDEGGTFAVEPM	457
QY	424	VKGLEETAVNDLISPPSGAPLISNADLGGFTHADVTAVPSTIFTSVPVFTGTSDVYAEAL	483

Db 458 VRGLBETAVNDLASPPSGAPLLSALGDVGGFPHRTSLTEVPSSMMYSPNFTSTTSIDFAET 517
 QY 484 NPSIIVRAGSFPPSQPNDRHVAFTDGGKWNFGQSEPGVTTGGTVAASADGSRFVWAP 543
 Db 518 KDVVVRAGNLD--SGP---HIAFTDNGANWFGTDPGSGVGGTVAAGADGSRFVWSP 572
 QY 544 GDPQGPVVYAVGFGNSWAAASQGVPAQAIRSDRVNPKTFYALNSGTFYRSTGGVTPQPV 603
 Db 573 --EGAGVQVTTGFTSWQASTGLPAGATVESDRVNPAITPYGKGRFYVSTDCGATFTAS 630
 QY 604 AA-GLPSSGAVGMFHAPGKEDLWLAASS-----GLYHSTNGSSWSAITGVSSAVNV 657
 Db 631 AATGLPAGD--GVRKALPFGGEGDVLGAGAADGYPGLWHSTDDGGTFTRLPGVDAADTV 688
 QY 658 GFGLKAPGSSYPAVFGVGTGAYRSDDCGTTVNLINDDOHOYGNWQQAITGDHANL 717
 Db 689 GFGLKAPGASQYTLTFAEIGVGRGIFRSTDAGATWTRVNDDAHQWGTGAITGDPRVY 748
 QY 718 RRVYIGTNGRGIVYGDII-----CGAPSG 740
 Db 749 GRVYVATNGRGIVYGDISTDGGGTDPG 775

RESULT 2
 D97013
 probably secreted sialidase, several ASP-boxes and dockerin domain [imported] - Clostrid
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
 C:Accession: D97013
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: D97013
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-839 <CUR>
 A:Cross-references: UNIPROT:097KK0; GB:AE001437; PIDN:AAK78895.1; PID:gi15023820; GSPDB:G
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC0919

Query Match 49.8%; Score 2009; DB 2; Length 839;
 Best Local Similarity 50.9%; Pred. No. 7.8e-110;
 Matches 377; Conservative 107; Mismatches 240; Indels 16; Gaps 12;

QY 3 TQPYTWSNVAIGGGFVDGIVFNEGAPGILYVTRDIDGMRYRWDAAANGRWIPLLDVWGN 62
 Db 37 SQGYKWDNAKIGAGGVVPAVFNKTEKDLIYARTDMGAYRWKANNKWIPIID--GFS 94
 QY 63 WGYNGVSVTAADPINTKWAAGVMTNSWDPNDAAILRSSDQATWQITPLPFLKGNM 122
 Db 95 WMLGCESTADPIDTNRVYIAGLYTNDQDENAYILSSQDKGNTKRYQLPFFKVGNN 154
 QY 123 PGRGGERLAVDPNDNIIYFCAPSGKGLWRSTDGATWSQMTNPDVGTIYANTDTITG 182
 Db 155 PGRNGERLQIDPNDKILYLARGNGLWKSSEDYGTWSKVDNPDPTGDYVQDPQNE-- 212
 QY 183 YQSDIQGVVVAADFSSSLGQASKTIFVGVAD--PNNPVFWRDGGATWQAVPGAPTGI 241
 Db 213 YPADKGVVWVEITDPTGKSPQTQMYVGAADKGNIIYVNDGGKTSWAKVGKPGYL 272
 QY 242 PHKGVPDPNVHLYIATNSNTGPGYDGSQGVWKFVSTGWTWTRISPVSTDTANDYFGVS 301
 Db 273 PHHGIL-ASDGLMYISYNTCPYDGSQGVWKFVSTGWTWTRISPVSTDTANDYFGVS 328
 QY 302 GLTIDRQHPNTIMVATQISWEDTIIFRSTDCGATWTRIMWTSYVNSLRVLVDI SAEP 361
 Db 329 GISVDAQNPNNVAVATLNRWPDDEIYRSTDAKWTWKPIWDMWNGYPNRTLGYNLDSAQ 388
 QY 362 WLTFFG-VQNPVPSPKLCWMDMAIDPNSDRMLYGTGATLYATNLDLTKWDSGQQIH 420

Db 389 WLDWCKTGVTTPDPLVLKLGMMMGDLLEIDPFNSDRMPYGTGATLYGTDDLTNNWDKGNVDI 448
 QY 421 APMWKLBETAVNDLISPPSGAPLLSALGDGLGFTHADVTAVPSTFTSPVFTTGSVDY 480
 Db 449 SYKANGIEBECANDVVPVTKGAQLLSAVGDDCGFFYHDDITKVPKXMTTTFNFSATSIDY 508
 QY 481 AELNPSIIVRAGSFPPSQPNDRHVAFTDGGKWNFGQSEPGVTTGGTVAASADGSRF 539
 Db 509 AESVFNFRVVRGVNVDTSKNQDKCGISYDGGKWNFSAGNSISGVYKAGTVAAGADAKTI 568
 QY 540 WMAPGDPQCPVVYAVGFGNSWAAASQGVPAQAIRSDRVNPKTFYALNSGTFYRSTGGV 599
 Db 569 WNSP--EEGANAISYTDNGNKMTPCGLPQGAIRSDRVNPKTFYALNSGTFYRSTDA 627
 QY 600 F-QPVAAGLPSGAGVGMFHAPGKEDLWLA--ASSGLYHSTNGSSWSAITGVSSAVNV 657
 Db 628 FTQSSQTLPTKKG--GI--PKTVIGHGDIWIAGGKDLWHSTDSGATFTKVGSDASD 685
 QY 658 GFGLKAPGSSYPAVFGVGTGAYRSDDCGTTVNLINDDOHOYGNWQQAITGDHANL 717
 Db 686 GLGKSKTDDGYPALYMDATIDGTAGIFRSDDDEGATWTRVNDDAHQYGSPTYCITGDPNKY 745
 QY 718 RRVYIGTNGRGIVYGDIGA 737
 Db 746 GRVYVATNGRGIVYGDIDGS 765

RESULT 3
 T00349
 Avicelase III - Aspergillus aculeatus
 C:Species: Aspergillus aculeatus
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
 C:Accession: T00349
 R:Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.
 submitted to the EMBL Data Library, June 1998
 A:Description: Avicelase III from Aspergillus aculeatus.
 A:Reference number: Z14141
 A:Accession: T00349
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-856 <ARA>
 A:Cross-references: UNIPROT:074170; EMBL:AB015511; NID:dl199887; PID:d1029971
 C:Genetics:
 A:Gene: aviii
 F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 41.6%; Score 1680; DB 2; Length 856;
 Best Local Similarity 46.2%; Pred. No. 1.3e-90;
 Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 1 ATTQPYTWSNVAI--GGGGFVDGIVFNEGAPGILYVTRDIDGMRYRWDAAANGRWIPLLDVW 59
 Db 21 AASQAYTWKVVVVGSGGGFTPGIVFNPASAKGAYARTDIDGGAYRLN--SDTWTPLMDW 79
 QY 60 ---WNNWGVGVVSTAADPINTKWAAGVMTNSWDPNDAAILRSSDQATWQITPLP 116
 Db 80 NDTWHDW---GIDALATDPVDTRVYVAVGMYTNEWDPNVGSILRSTDQGDITWETKLP 136
 QY 117 KLGNMPPGEMGERLAVDPNDNIIYFCAPSGKGLWRSTDGATWSQMTNPDVGTIYAN 176
 Db 137 KYGVNMPGRGEMGERLAVDPNKNSILYFGARSGHGLWKSTDYATGATWSNVTSTFTW 196
 QY 177 PTDITTYQSDIQGVVVAADFSSSLGQASKTIFVGVADPNNPVFWRDGGATWQAVPGA 236
 Db 197 SSST--YTSDFVGIAWVTFDSTSGSGSATPIFVGADAGKSVFKSEADAGTAWWSGE 254
 QY 237 PT-GFIPKHGVPDPNVHLYIATNSNTGPGYDGSQGVWKFVSTGWTWTRISPVSTDTAN 295
 Db 255 PQYGFPLPHKGVLSPEKTYLISYANGAGPYDGTNGTVHKYNTISGVWTDISP----TSLAS 311
 QY 296 DYFGYSGLTIDRQHPNTIMVATQISWEDTIIFRSTDCGATWTRIMWTSYVNSLRVYL 355
 Db 312 TTYGYGGLSVDLQVFGTLVVAALNCWMPDELIFRSTDSGATWSPIWEMNGYPSINYYSY 371

356 DISAEPWLTFFGVQPNP - PVPSPKLGWMDZAMAIIDPFNSDRMLYGTGATLYATNDLTWKWDS 414
 372 DISNAPWIDTSTDOFPV --RVGMVFEALAIIDPFDSNNHMLYGTGLTVYGGHDLTNWDS 428
 415 GGOIHAPVMKLEETAVNDLISPSGAPLIISALGDLGGFTHADVTAVPSTIFTSPVFTT 474
 429 KHNVTVKSLAVGTEEMAVLGLITPPGCPALLSAGVDDGGFYHSDLDAAFPANQAYHTPTTGT 488
 475 GTSVDYAEALNPSIIVRAGSFDPSSQPNDRHVAFTSTDDGKNWFGQSGPGVTTGGTVAASA 534
 489 TNGIDYAGNKPNI VRSGASD --DYPT--LALUSNFGSTWTYADYAAASTSTGTGAVALSA 543
 535 DGSRFWAPGDCQPVVYAVGFGNSWAASQGVPANAIQRSDRVNPKTFYALSGNTGTFYRST 594
 544 DGTVLIMSTSGALVSKSQG--TLTAVSSLPAGVIAASDKSDNTVFYGGAGAIYVSK 600
 595 DGGVTQPVAAGLPPSSGAVGVWFMHAPVKGEDLWLAASSGLYHSTNGGSSWSAI -TGVSS 653
 601 NTATSPFTKVS -LGSSTTNAI -RAHPSIAGDVMASTDKGLWHSTDYDSTFTTQIGSGVTA 658
 654 ANNVGKCAPGSSYPAVFTVGTIGVTGAYRSDCGTTTWLINDDQHOYGNWGOA -IT 711
 659 GWSFGFGKASSTGSYVVIYFFTTIDAAGLKFSEADGTWQVVISDASHRFGS -GSANVVN 717
 712 GDHANLRRVYIGTNGRGIVGDI GGAPSG 740
 718 GDLQTVGRVFRGHERPGHLLRQSQREPAG 746

RESULT 4
 F72393
 hypothetical protein TM0305 - Thermotoga maritima (strain MSB8)
 C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
 C:Accession: F72393
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D. A.
 C.M.

Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A:Reference number: A72200; PMID:99287316; PMID:10360571
 A:Accession: F72393
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-707 <ARN>
 A:Cross-references: UNIPROT:Q9WYE1; GB:AE001712; GB:AE000512; NID:q4980799; PIDN:AAD3539
 A:Experimental source: strain MSB8
 C:Genetics:
 A:Gene: TM0305

Query Match 28.8%; Score 1160.5; DB 2; Length 707;
 Best Local Similarity 35.2%; Pred. No. 2.4e-60;
 Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;

QY 6 YTSNVAIGGGFYDGVIFVNEGAPGLYVTRDIGMYRWDAANGRWIPLLDVGVNNGWY 65
 21 FEWKSVEINGGFPVGLIIFHPASGLIYARTDVGLYRWDEETKWKQLFDLRRDQSDY 80
 66 NGVVSIAADPINTNKVAAVGMVYTNSSWDPNPDGAILRSSDQCATWQITPLP ---FKLGGNM 122
 81 MGVLVALDPSPDKRIYAMTKYQTDW -AGYGAILISEDYGETWTI VNLDKYGIKVGNE 139
 123 PGRGWRGLAVDPNDNILLFVGPAGSKGLWRSTDSGTATWQMTNFPDVGVIANTPTDTTG 182
 140 DGRNAGERLQVDPNPFSSVFMGT -TKYGLMKSEDFGKNWKKVSPFST----- 186
 183 YQSDIQGVVVAFOKSSSSLSQASKTI FVGVAADPNPNVFWSRDGGATWQAVPGAPTGP 242
 187 -----SVTFVLFEKSGEKGSPTRIFVCGSEPKG -IFVTEDCGITWNLVNLPLNDLIP 239
 243 HKGVFDPVNHVLYIATNTGNGPFGDGGSDVWVKFVSTGWTTRISPVPSSTDTANDYFGVSG 302

Db	1215	NGTVVATATPTGNTGQAAATTVDAVAPPAPVIDPNSGTTISCTAAGAKVILTDGNGN	1274
Qy	219	PV-FWSDGGATQAVPCAP-----TG-----FIPKHGVFDPVN	251
Db	1275	PIGTTADGSGNWSFTPTPLANGTVNNAQDPAGNTGPGSTVDVAVAPNTVVPNSN	1334
Qy	252	HVLVIATSNTPGYP---DGSSGDYWKFSVT-SGTW---TRISFPVS---TDTANDYFGYS	301
Db	1335	GNLLNGTAEPGSTVTLTDGNGNPIGQTTADGSGNWSFTPGSQLNGTVNVTASDAAGNT	1394
Qy	302	GL-----TIDROHPNTIMY---ATQISWNP---TILFRSTDPG---ATWTRIWD	342
Db	1395	SLPATTTVDSSLPSPQVDPNSGVSISCTADAGNTIIII--TDGNGNPIGQVTDAGSGNWS	1452
Qy	343	WT-----SYPNRSRLRYLVIDIAEPWL---TFGVQNPVPVPSKLGWMDAMALDPNS	392
Db	1453	FTPIGILPDGTVVNVVAKSPNSVDSAPAVITVDGVAPAPV-----IDPSNG	1499
Qy	393	DRMLYGT---GATLYATNDLTKWDSGGQIHIAPVMVKGLEETAVNDLISP---PSGAPLI	445
Db	1500	TE-ISGTAEAGATVILT-----DGGN-----PIQATADGSGNWTTPSTPLANGTVIN	1548
Qy	446	SALGDLGGFTH-----ADVTAVPSTIP--TSPVFTTGTSTVDYAEIENPSIIVRAGSFDPS	497
Db	1549	AVAQDPAGNTSGPASVTVDALAPPAPVINPNSGVVISGT---AEAGATVILTDGNGNPI	1604
Qy	498	SOPNDRHVAFTDCGKNWF-----QGSERPGVT-----	525
Db	1605	GQ-----VTADGSGNWSFTPTGTPLANGSVINALAQAAGNNSPTSATVDSLAPAPV	1657
Qy	526	-----TGTVAAADGSRFVWAPGDPGPV---VYAVFGNSWAAQGVPPANAQIRSDRVNP	579
Db	1658	IDPNSGVIACTAAGATVILTLDGNGNPIGQVTDAGSGN-WSFTPTGP-----	1704
Qy	580	KTFVALNGTFYR-----STDGVV--TFQVVAAGLP---SSGAVGVWFHAPVPGKGD	626
Db	1705	-----LSNGTVVNAVAQDAAGNTSGPSTTVDAVAPATFVIDPNSGVLSGTAEPGRVI	1759
Qy	627	LWLAASSGLYHS--TNGSSWSAITGV-----SSAVNVFGKSAPOSSYPAPVVGITIGGVT	681
Db	1760	LTDGNGNPIGQTLADGSGNWSFTPTPLANGTVNNAVAQDPAGNTSGPASFTVDTVPAT	1819
Qy	682	GAYRSDDCGTTWVLINDDQYGNWGOAITGDHANLRVYIGTNGRGIVVGDIGGAPSG	740
Db	1820	-----PVINPSN-----GSVITGT-AEVGAKVILTDGNGNPIGETTADGSG	1859
RESULT 6			
B95965			
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag			
C:Species: Sinorhizobium meliloti			
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004			
C:Accession: B95965			
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan			
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001			
A:Title: The complete sequence of the 1,883-kb pSymb megaplasmid from the N2-fixing endo			
A:Reference number: A95842; MUID:21396508; PMID:11481431			
A:Accession: B95965			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-2174 <KUR>			
A:Cross-references: UNIPROT:Q92U08; GB:ALU591985; PIDN:CAC49389.1; PID:gl5140875; GSPDB:C			
A:Experimental source: strain 1021, megaplasmid pSymb			
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,			
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;			
L.; Hyman, R.W.; Jones, T.			
Science 293, 668-672, 2001			
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,			
hebaault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.			
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.			
A:Reference number: A96039; MUID:21368234; PMID:11474104			
A:Contents: annotation			

C:Genetics:			
A:Gene: SMD21548			
A:Genome: plasmid			
Query Match			
Best Local Similarity 5.0%; Score 202.5; DB 2; Length 2174;			
Matches 176; Conservative 90; Mismatches 312; Indels 197; Gaps 36;			
Qy	12	AIGGGGVDGIVFNEGAPGILYVRT-DIGGMRYRMDAANGRWIPLLLDVGMNNGVGVVS	70
Db	976	SIGGG-----GNGGVGSNSKEIGSGFNLT-----NVGVGSGS	1011
Qy	71	IAADPINTNKVAAVGMVYTNNDNDGAILRSSDQGA-----TWQITPLPK	117
Db	1012	GGASGN-----AIVGL-----DSGTHLQTSGSGARGVIVQSIGGGGTSQGAASVGLS	1059
Qy	118	LGMNPGMGMERLAVDPNNDNLLYFGAPSGKGLWRSTDSGATWSQMTNFPDVGIIANP	177
Db	1060	ASASLPGGG-EEAAEABEESBEGSAFSASVGVSGRTGGSGS-SGTVNVTTAGT-----	1112
Qy	178	TDTTGYOSDIOGVVWVAFDKS---SSSLGQASKTIFVGVADPNPNVFWSRDG-----	226
Db	1113	--ISTFGADADGVLAQSIGGGGLGSGVQAASSGSEPLDDGDESEACASGNGDDHGY	1170
Qy	227	-----GATWQAVPCAPTGFIPKHGVFDPVNVHLYIATS-----NTGGPYDGSSG	270
Db	1171	CFGVSVGAT---IDGGTGGTAAANGNAVTLTHAGHATAGDWADGIQAQSIGG---GGGAG	1225
Qy	271	DWKFSTVSGTWTTRISPVSTDDTANDYFGYSLGTLIDROHPNTIMVATQISWWPDTII	330
Db	1226	GTSTASGQATANTIVGVGGSGAGGAGGAVGIIITFDNHNHSISITAGYSAY---GVLLQS	1282
Qy	331	TDGGATWTRIWDWTSYPNRSRLRYLVD-----ISABPWLITFGVQPNPPVPS	375
Db	1283	IGGGG-----QGGDGSDEAAGRITVGGFGSGSGAGSGGWATAKGNINLSTSG-----	1332
Qy	376	PKLGWMDAMAI--DPFNSDRMLYGTGATLYA-----TNDLTKWDSGGQIHIAPMVKGL	427
Db	1333	-----DDAHGIVAQSIGGGGVGGAGSSTAEEKHSHTIDLIVGGSG-----VGS	1382
Qy	428	BETAVNDLISPPSGAPLISAL-----GDLGGFTHADVTAVPSTIFTSTPVTGTSVDY	480
Db	1383	VDLSVGTTLT--TSGARALGLVAQSIGGGGIGGAGEAD--SIASLVVGG---SGGTTIDG	1436
Qy	481	AELNPSIIVRAGSFPSSQPNDRH--VAFSTDTGKN-----WFGSPSPGGVTTGG	528
Db	1437	GAVTVDLTSQSSI---TTQGIAAHGLVAQSIGGGGVGGAGSAGAPLSTGNSPGSYDGG	1493
Qy	529	TVAASADGSRFV---WAPGDPQPVVYAVFGNSWAAQGVPPANAQIRSDRVNPKTFYAL	585
Db	1494	DVAVTADGSIPTRGDYAFGLVAQSIGGGGGFGGN-ATSAFIGSNGNLSSDG-----	1543
Qy	586	SNGTFFRSTDGGVTFQPVV-----AGLPSSGAVGVWFH-AVPGKEGD-----LWL	629
Db	1544	KSGNVTVSLDAGRTIQAGSKDSIGIFAQSDAGTDNNGTIDVTNGTIVTGTGSGDNGAGI	1603
Qy	630	AASGLYHSTGSGSWSAITGVSSANVNVFGKSAPOSSYPAPVAVVGTI--GGVTGA	683
Db	1604	SAGKDNVTVNNSGGNVAASGV--AVQVTAGNNSPEDSTLVVNNAGTISGSVKGA	1656
RESULT 7			
S75251			
hypothetical protein slr1028 - Synecocystis sp. (strain PCC 6803)			
C:Species: Synecocystis sp.			
A:Variety: PCC 6803			
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004			
C:Accession: S75251			
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;			
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda			
DNA Res. 3, 109-136, 1996			
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis			
S.			

QY 516 -----FOGSPGGVTTGGTVAASA-----DGSRF-----VWAPG----- 544
Db 2929 LRANQSGSPRFAIDGSPNSQAGIALNGGDLNGDFADFIIGAFENNLQYNQIIVF 2988
QY 545 -----DPGQPVVAVG-----FGNSWAASQGVPAQAIRSDRVN-----PKTFVAL 585
Db 2989 IENGELSDDDKYSYILYLDGNQTTQMGSGDQWQANQ-VMTN-QVATNWNNSRRPPEAVITGQ 3046
QY 586 SNGTFYRSTGGVTFO-----PVAAGLPSSGAVGVMHAPGK 623
Db 3047 SNGDIWTPYPGNQWQSGWGLPAEINELAVNMTSGNPIITAGLGKG--GIEYY----- 3099
QY 624 EGDWLAAASGLYHSTGGSSW-SAITGVSSAVNVGFKSAPSYPAVFVVGTTGGVTG 682
Db 3100 NGSTWV--NNGPYQ-----GDGWSRAITQMA-----VQWGED--GS--PSQIVVGLDAGVI 3145
QY 683 AYRSDDCGTTWVLINDQHQYGNWQAITG-----DHANLRVYIGTN 725
Db 3146 YVNTQ-----SGWRTIN-----NFGKSVTQLSVQWQBASNPNIVVGLDSEV-QVYQGSN 3194
QY 726 GRGIVYGDIG 735
Db 3195 GVMTQFHDG 3204

RESULT 9
T13256
tail-host specificity protein homolog - Lactococcus lactis phage BK5-T
C;Species: Lactococcus lactis phage BK5-T
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: T13256
R;Boyer, J.D.; Davidson, B.E.; Hillier, A.J.
Appl. Environ. Microbiol. 61, 4089-4098, 1995
A;Title: Sequence analysis of the temperate Lactococcus lactis bacteriophage BK5-T and
A;Reference number: Z17646; MUID:96064422; PMID:8526523
A;Accession: T13256
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1904 <BOY>
A;Cross-references: UNIPROT:Q38319; EMBL:L44593; NID:g928826; PID:g928828; PIDN:AAA98579

Query Match 4.5%; Score 182; DB 2; Length 1904;
Best Local Similarity 20.7%; Pred. No. 0.013;
Matches 189; Conservative 83; Mismatches 260; Indels 380; Gaps 55;

QY 1 ATTOPYTWNSVAICGGGFVDGIVFNEGAPG-----ILYVRTDIGMYRWDANGR 50
Db 838 AKPSDYTWS--LIRGNDGKGCATGKGVAGKGVGKTTVTIYALSSSG-----TDKPTNG 891
QY 51 W---IPLLDWGVNWNWYGVVSTAADPINTNKWAAVGMVNTNSWDNDGAILRSDQGA 107
Db 892 WTSQVPTL-----VKGQYLTWTKVTM-----YDTS-----SSETG- 921
QY 108 TWQITLPLPFLKGNMPCRMGERLAVDPNNDNILYFGAPSGKGLWRSTDGATW--SOMTN 166
Db 922 -YSYIYI--AKDGNNGDGIAGKGVGKTTIYAVGTSG-----TTAPASGNSQVFN 973
QY 167 FP-----DVGTIYANPTDTTQSDIQGVVWVAFDKSSSSSLGQASKTIFVGVADPN 218
Db 974 VPAGQFLWTKTVIYTDN--TSETGYSVAMGVKGDGDPGNNTN-----GIA----- 1020
QY 219 PVFWSRDG-----GATWQAVPG---APTG---FIP---HGKVP-----DP 249
Db 1021 ----GKDGKIGKATAITYQASPNGTTAPTGTWSASVPPVAKGSFLWTRTIWYTDNTTET 1076
QY 250 VNHVLYTATSWTGGPYGSSGVKFSVTSCTWTRISPVSTDTANDYFGYSLTIDRQH 309
Db 1077 GYAVAYMGNTGNG--HDGFGP-----KDGTKITTTIT-----YAGSTSGTTP 1118
QY 310 PNTIMVATQISWPPDTTIFRSTGATWTR-IWDWTSYPNRSRYVLDISAEPMLTFCVQ 368
Db 1119 EN-----NGWTSTVP-TVABGNLYWTKVIWYTD--NTS----- 1149

QY 369 PNPPVPSKLGWDEAMAI-----DPFNSDRMLYGTGATLYATNOLDTKWDSGQIHAPM 423
Db 1150 -----ETGYSVAMGVKGDGKDPGNN-----GTNGIAGKDG----- 1180
QY 424 VKGLEETAVNDLIIP-PSGAPLISALGDLGGFTHADVTAVP--STIFTSFVFT----- 473
Db 1181 -KGIKATAITYQASPNGTTAPT-----GTWSASVPPVAKGSFLWTRTIWYTDNTTTE 1231
QY 474 TGTSVDVYAEIN-----PSIIVRAGSPDPSPQPNDRHRVAFSTDGKKNWF 516
Db 1232 TGYAVAYMGNTGNGCHDGPCKDGTGKITTITTYAGSTSGTTPNN-----GW- 1279
QY 517 QGSEBPGVTTGGTVAASADG---SRFVW-----APGDPGQPVVYAVG 555
Db 1280 -----TSTVPTVAEGNVLTKTVWYTDNTSETGYSVAMGVKGDGKDP----- 1323
QY 556 FGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTTFYRSTGDTGVTQPVAAAG----- 606
Db 1324 -GNN--GTNGI-AGKDGKIGKATAITYQASPNGTTAPTGTWSASVPPVAKGSFLWTRTIW 1379
QY 607 -----LPSSGAVGVN-----PHAVPKEGD-----LWLAASSGLYHSTNGSS-- 644
Db 1380 TYTDNTTETGYAVAYMGNTGNGNNGHDGFEKGTGTGKTTITTYAGSTSGTTPNNGWISTV 1439
QY 645 -----W-----SAITGVSSA-VNVGFKSAPGSYPVAVFVVGTTGGVTG-- 682
Db 1440 PTVAEGNVLTKTVWYTDNTSETGYSVAMGVKGDGKDPGNN-----GT-NGIAGKD 1491
QY 683 -----AYRSDDCGTTWVLINDQHQYGNWQAIAI-----TGDHANLRV----- 720
Db 1492 GKGIGKATAITYQASPNGTT-----APTGTWSASVPPVAKGSFLWTRTIWYTDNTTTE 1543
QY 721 -----YIGTNG 726
Db 1544 TGYAVAYMGNTG 1555

RESULT 10
F70846
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: F70846
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70846
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2523 <COL>
A;Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17115.1; PID:e125111
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: PPE

Query Match 4.3%; Score 175.5; DB 2; Length 2523;
Best Local Similarity 21.8%; Pred. No. 0.047;
Matches 175; Conservative 92; Mismatches 327; Indels 209; Gaps 38;

QY 10 NVAIGGGFVDGIVFNEGA--PGIYVTRDIGMYRWDAAAN-GRWITPLLDWGVNNGWYN 66
Db 1522 NTGICNSGIASGLFNAGGNTGVVAGSYNTGSGFNAGQANTGGFNPGSVNTGLNTG-- 1579
QY 67 GVWSTAADPINTNKWAAVGMVNTNSWDNDGAILRSS-DQATWQITPLPFLKGNMGP- 124
Db 1580 -----DINTG-----VANSGDVNTGTAFISGNYNGAFWR-----GDYQGL 1614
QY 125 RCMGERLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNPPDVGTYIANPTDTTGYQ 184

1615 LGFSRPAVLQ-----TPFLDLTLTGLGSVWIPADIPAIR 1652
 185 SDIQGVVWV-APDKSSSLGQ---ASKTIFVGVAD-----PNNPVFWRDGGATWQA 232
 1653 PEFSANVAIDSFVPSIPQIDLAATTVSVGLGPIITVPHLIDIPRPVPTLNYLFQSG--- 1709
 233 VPGAPGPIPHKGVDPVNHVLYIATNTGPGYDGSVDVWKFVSTGVTWTRISVPSTD 292
 1710 -PGGPKIGTITGLFTPIGTLPTLALSIQIVIGASSQGTITAFLANLPSTVPTVIDEIP 1768
 293 TANDYFYS-----GLTIDROHPNTI-----MVATQISWMPDTIIFRSTDDG 334
 1769 LLASITGSEPDVIFPGGLTIPAMNPLSINLSCGTGAVTIPAITIGETIPFDLVAHSTLIG- 1827
 335 ATWTRIMDWTSPNRSRLRYLDISABPMLTFGVQVNPVPKLGWMD----- 383
 1828 -----PVHILIDLPAVP--GFG-----NTGAPSSGFFNSGAGGVSGFGNV 1866
 384 -AMADPFN--SDRMILYGTGATLYATNDLTKWDSGQIHIAPMVKLEETAVNDLISPSPG 441
 1867 GAMVSGGNQAPSALLGGSGGVFNAGTL-----HSGVLNFGSGMGLFNTSVLGL-----G 1917
 442 AP-LISALGDLGFTHADVTAVPSTIFTSPVTTGTSVDYAEINPSIIVRAGSFDPSSOP 500
 1918 APALVSGLGSGVQ-----QLSGLLASGTA-----LHQVLNFGGLADVGL-- 1957
 501 NDRHAFSTDDGKNVFGQSEPGVTTGGTVAASADGRFV-WAP-----GDPQPVVY 552
 1958 --GNVGLNVGDFNLGAGNVGFGVGGNIGNNVGLGVNGFGLGNSGLTGLMGLG 2015
 553 AVFGNSWAASGVANQIRSDRVNPKTFYALSNGTF-----YRSTDDGV 598
 2016 NTFGNGAGSYNFGI-ANM-----GVNIGCFANTGSGNFGIGLTDNLGTGFGGFGTSGNV 2069
 599 TQPVAAGLPSSGAVGV-MFHAVPGKEGDLWLAASSGLVHSTNGSSSAITGVSSAVNV 657
 2070 -----GLFNSGTVNGVFNFGTGN-----WGVFNSSGY-NTGIGNSGIASTGLFNAGGF 2117
 658 GFGKAPSSSYPAVFGVGTGAYRSDDCGTTWVLINDQHOYGNWQA-----ITG 712
 2118 NTGVNAGSYNTGTFNAGAN--TGFPNPGSVNTGWLNTGDTINTGVNSGDVNTGAFISG 2175
 713 DHANLRRVYIGTNGRGIYVDIG 735
 2176 NYSN-----GAFWRGDYQGLLG 2192

RESULT 11
 AE1905
 outer membrane secretion protein alr0791 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
 C;Accession: AB1905
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
 A;Reference number: AB1807; MUID:21595285; PMID:11759840
 A;Accession: AE1905
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-993 <KUR>
 A;Cross-references: UNIPROT:Q8YQ05; GB:BA000019; PIDN:BA072748.1; PID:g17130136; GSPDB:G17130136
 A;Experimental source: strain PCC 7120
 C;Genetics:
 A;Gene: alr0791

Query Match 4.3%; Score 172; DB 2; Length 993;
 Best Local Similarity 19.6%; Pred. No. 0.022;
 Matches 168; Conservative 96; Mismatches 325; Indels 270; Gaps 44;

45 DAAGRWM-IPLLDWVGNWNNVNGVSVIAADPINTN-KV-----WAAVGMTNSW 92
 58 DMSNGTWEVTLNNGTSWTTFGTPTSLTAARLLPSNANTKIRFHPNANFSGTADINFY--AW 115
 93 DPNDG-----ALRSSDQATWQITPL-----PFLKGN-MPG----- 124
 116 DQTTGTSSTANILAGKGTAFSTNYEGASITVTPVNDTAPTLANAGSTPLPINEDAPL 175
 125 -RGMGERLA-----VDPNNDNILYFCAPSGKGLMR--STDSGATWSQMTNFPDVGTY 173
 176 ISNKGSLIADLVRLGISDSDDPYQIAVAGADNNGSWQYSLDGGVNLNFGNASDNSAT 235
 174 IANPTDTTGYQSDIOGVWVMAFCKSSSSSQASKTIFVGVADPNPNNVFWVRDGGATWQAV 233
 236 LLLPSIR-----LYDGSLSLPTSQWMLKFGASPPVPLF-----PVGQTSLI 279
 234 PGA-----PTGFTPHKGVDPVNHVLYIATNTGPGYDGSVDVWKFVSTGVTWTRISPVPS 290
 280 TGGIQLNSSSISGSG--YSNYSYAPILFQAPELDPVKGTTFISFDVKINGETHSTSD-DN 337
 291 TDTANDYFCYSGLTIDRQHPNTIMVATQISWMPDTI-----IFRSTDDG 334
 338 GGIQDRAGFSVIVTSDTKAI-----ELGFTDEIWAQTASPLFTHSTITERAFNTTAA 393
 335 ATWTRIMDWTSPNRSRLRYLDISABPMLTFGVQVNPV-----PSPK 377
 394 VT-----RVHLVVENNTYKLFAPDSSPTILSGNLRDYSAFNHSTAAPSP- 437
 378 LGMWDEAMAIIDPFNSDRMLY-GTGAT-LYATNDLTK----- 411
 438 ---ITSPLFPDVFETPNFLGDNLTSAQASSNLTQVELQTNTRVRFPVNADYNGQANLT 493
 412 ---WD-----SGQIHIAPMVKLEETAVNDLISPSPGAPLISALDGLGFTHADVTAVP 463
 494 FRAWGSGNVAGG-----TTGVNAAVN-----GNATFSSNTLTA-- 528
 464 STIFTSVP-----FTTGTSDVYAEINPSII-----VRAGSFDPSSQPNDRHVAFS 508
 529 -SITVSPINNPITQTTGLDKLYGTANEDIIINGEGNDYLFGRAGNDTLDGEGNDYLFPG 587
 509 T-----DGGK--NWFGQSEPGVTTGCTVAASADGRFVWAPGDPGVVYAVGVFNSW- 560
 588 TGNLTDGEGSDLLYGNEDNDIINGGVGNDLDG-----GTGDD-----ILRGGTGNDIY 638
 561 ---AASQGVANQIRSDRVNPKTFYA-----LSNGTFYRST--DGGVTFQFVAAGLPSSG 611
 639 TVDTVGDVTEENPNEGTDKNVSYISWTLGANLENLTLLGNTIIDG--TGNELDNHIIGNN 696
 612 AVGVMFHAVPGKEGDLWLAASSG--LVHSTNGSSWSAITGVSSAVNVGPKSAPGSYP 669
 697 AV-----NRLEGGDNDWILGKQNDILIGGNDRLNGETG-EDTLEGGLGND----- 744
 670 AVFVVGTTGGVGTGAYRSDDCGTTWVLINDQHOYGNW-----GQAITGDHAN 716
 745 -VVEIDSVGV--IEAADAGIDTVISSVD-----WTLGVLENLTVLGVNQATLIGND 795
 717 LRRVYIGTNGRGIYVDIG 735
 796 LDNRITGNNAADNVLFGEAG 814

RESULT 12
 AE2254
 hypothetical protein alr3588 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
 C;Accession: AB2254
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
 A;Reference number: AB1807; MUID:21595285; PMID:11759840

QY 642 -GSSNSALTGVSSAVNVGFGKSPAGSYPVAVFVGTGGT-----GAYR-----SDDCG 690
 Db 4776 TQGSY-LYTNVSNASN--WTQSALGGSQS-----GTTSGVAAMDYDWDGAVDVLVSKQSG 4827
 QY 691 TTWVLINDDOHQYGNWQAITGDHANLRVYVGTNGRGIVYGD 733
 Db 4828 SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVIYGN 4860

RESULT 14
 F90696
 hypochelical protein EC0542 [imported] - Escherichia coli (strain O157:H7, substrain R1)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
 C:Accession: F90696
 R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: F90696
 A:Status: preliminary
 A:Molecule type: DNA
 A:Cross-references: UNIPROT:O8X2T1; GB:BA000007; PIDN:BA033965.1; PID:gl3360000; GSPDB:G
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: EC0542

Query Match 4.2%; Score 168.5; DB 2; Length 5291;
 Best Local Similarity 20.1%; Pred. No. 0.32;
 Matches 190; Conservative 82; Mismatches 270; Indels 401; Gaps 46;

QY 27 GAPGILYVRTDIGMYRMDAANGRWIPLLDWGMNNGVYVIAADPINTNKVAAVG 86
 Db 4186 GANFPAQISTDNGA-----TWVN-----VTVAADSLN-----4212

QY 87 MYTNSWDNDGAILRSSDQATWQITPLPKLGGNPGRMGERLAVDPNN-----137
 Db 4213 -----WSYVDGRTLTNGT--TTQVRRV--DLAGNV-GATSSQSALIDTVNPAQVLTIAS 4262
 QY 138 -----DNILYF-----GAPSKGLWRSTDGATWSQNTNPDVGTYIAN 176
 Db 4263 ISTDGSSATPITSDTMTLTGSLGAGLAGEVAQISLDGATWTLTITNGTQMTYTD 4322
 QY 177 PDDTTG---YQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPN---NPVFWSRDGGATW 230
 Db 4323 RLTDGSGVYQVRV-----LDLAGNTGPPVSKTVVDVTINPTATPIVSVTDDVGQRQ 4375

QY 231 QAVPGAPTGFIPHKGVFDPVNRVLYIATSNTPGPPYDG-----SSGDV-----272
 Db 4376 GTLSSQ-----ATDDTTPLLNGVLSPALASGEVVYLYRNLGLGA 4416

QY 273 -----WKFQ---VTSQGTW---RISFVPSTDTANDYFGVSGLTIDRQHPNTIMVAT- 317
 Db 4417 VTMWALNWTYSDSGLVSGATYSARVVDLAGNITSSDF---VLTVDTSIPTTLAQITS 4473

QY 318 -----QISWN---PDT---325
 Db 4474 QTRDTTPIISGVITAAALASQYVEVINGKTYTSEPGAVVDPAHNTWYVQLPDTAL 4533

QY 326 -----IIFRSTDDGATWTRIDWTSYPNRSLRYVL 355
 Db 4534 TVSATAYTVTAQVKSSAGNNGNANTSGVTVNAALIDYPTPTASKTTAW---GLTYGL 4590

QY 356 DISAEFWLTFGVQPNPPVPSKLGWMDAMALDPFNSDRMLYGTGATLYATNDLTKWDSG 415
 Db 4591 D-SHGWTVLANQQVWQSDTP-LTWSKTALT-----LYQSGNN-YATSSIAIDYDRN 4638

QY 416 QQIHLAPWVKGLEETAVNDLISPPSCAPLIS---ALGDLAGGT-HADVTAVPSITFTSP 470
 Db 4639 G-----TGDLFITRDDYGTGYINGFTNNGDGT-----FSSA 4669

QY 471 V-FTTGTSDVYAEALNPSIIVRAGSPDPSSQPNDRHVAFST--DGGKNWFGQSEPG-----522
 Db 4670 IQVTGTLTWYGI-----VAFDKBGDGYLDFMIGDAGGPDSENT 4708

QY 523 -----GVTTCGTVAASADGSRFFWAPGDPQGVVYAVGFGNSWAASQGVPAQAIRSDR 576
 Db 4709 FLNNAGTLVGNSTTSNGSGSATVGG-----AVTGYLSLNGSGVDLNNDRIDL 4758

QY 577 V-----NPKTFVALS-----NGTFYRSTDGGVTFQPVAAAGLPSGAGVGMFHAVPQK-EG 625
 Db 4759 VQHTYNLNNYYTLSSLIQNGTFFWQNTTNTFLSGAGSGAMSSSVMTWADFDDGDDM 4818

QY 626 DLWLAASSG-----LYHSTNG-----641
 Db 4819 DLFLPASQGRANYGSLLENTFNTGVLGCPVAVAGATATTYASQFSLAVDWNHGLMDIARIAQ 4878

QY 642 -GSSWSALTGVSSAVNVGFGKSPAGSYPVAVFVGTGGT-----GAYR-----SDDCG 690
 Db 4879 TQGSY-LYTNVSNASN--WTQSALGGSQS-----GTTSGVAAMDYDWDGAVDVLVSKQSG 4930

QY 691 TTWVLINDDOHQYGNWQAITGDHANLRVYVGTNGRGIVYGD 733
 Db 4931 SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVIYGN 4963

RESULT 15
 A28452
 proteoglycan core protein precursor, cartilage - rat
 N:Alternate names: aggrecan
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 09-Jul-2004
 C:Accession: A92623; A28453; A28095; A28452
 R:Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
 J. Biol. Chem. 262, 17757-17767, 1987
 A:Title: Complete primary structure of the rat cartilage proteoglycan core protein deduced
 A:Reference number: A92623; MUID:88087070; PMID:3693370
 A:Accession: A92623
 A:Molecule type: mRNA
 A:Residues: 1-2124 <DO>
 A:Cross-references: UNIPROT:P07897
 R:Doerge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
 J. Biol. Chem. 263, 10040A, 1988
 A:Reference number: A30069
 A:Contents: annotation; revision to residue 698
 R:Doerge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.
 J. Biol. Chem. 261, 8108-8111, 1986
 A:Title: Partial cDNA sequence encoding a globular domain at the C terminus of the rat
 A:Reference number: A23835; MUID:86250698; PMID:2424893
 A:Accession: A23835
 A:Molecule type: mRNA
 A:Residues: 1856-2124 <DO2>
 A:Cross-references: GB:M13518; NID:G206104; PIDN:AAA41836.1; PID:G206105
 R:Neame, P.J.; Christner, J.E.; Baker, J.R.
 J. Biol. Chem. 262, 17768-17778, 1987
 A:Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan amino-ter
 A:Reference number: A28453; MUID:88087071; PMID:3693371
 A:Accession: A28453
 A:Molecule type: protein
 A:Residues: 20-37, 'w', 39-60, 'E', 62-64, 'X', 66-69; 70-83; 84, 89-148, 'L', 150-238, 'S', 240, 'A'
 C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; B
 C:Keywords: glycoprotein
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-2124/Product: proteoglycan core protein #status predicted <MAT>
 F:44-135/Domain: immunoglobulin homology <IMM>
 F:170-247/Domain: link protein repeat homology <LNK1>
 F:268-349/Domain: link protein repeat homology <LNK2>
 F:504-581/Domain: link protein repeat homology <LNK3>
 F:602-683/Domain: link protein repeat homology <LNK4>
 F:1914-2034/Domain: C-type lectin homology <LCH>
 F:2041-2097/Domain: complement factor H repeat homology <PHD>
 F:126, 239, 333, 387, 611, 667, 1842/Binding site: carbohydrate (Asn) (covalent) #status pred

Query Match 4.2%; Score 168; DB 2; Length 2124;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:44:57 ; Search time 92.0064 Seconds
(without alignments)
4118.611 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWTSNVAIGGGGFDV.....YIGTNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2429.5	60.2	882	Q82M04	Q82M04 streptomyc
2	2371	58.7	890	Q86727	Q86727 streptomyc
3	2062.5	51.1	996	Q9AQH0	Q9AQH0 caldicellul
4	2009	49.8	839	Q97KK0	Q97KK0 clostridium
5	1982	49.1	842	Q70DK5	Q70DK5 clostridium
6	1765.5	43.7	857	Q8TFP1	Q8TFP1 aspergillus
7	1753.5	43.4	806	Q9P4T8	Q9P4T8 agaricus bi
8	1680	41.6	856	Q74170	Q74170 aspergillus
9	1579	39.1	838	Q7Z9M8	Q7Z9M8 trichoderma
10	1521.5	37.7	861	Q7S1W6	Q7S1W6 neurospora
11	1405	34.8	739	Q82K30	Q82K30 streptomyc
12	1162.5	28.8	751	Q8PLM5	Q8PLM5 xanthomonas
13	1160.5	28.8	707	Q9WYEL	Q9WYEL thermotoga
14	1145	28.4	751	Q8EP9U5	Q8EP9U5 xanthomonas
15	1117	27.7	812	Q830D2	Q830D2 geotrichum
16	979	24.3	776	Q764N8	Q764N8 geotrichum
17	409	10.1	196	Q8J1H7	Q8J1H7 agaricus bi
18	368.5	9.1	1064	Q7X2U2	Q7X2U2 uncultured
19	325	8.1	911	Q72WN1	Q72WN1 desulfovibr
20	268.5	6.7	1109	Q7NME3	Q7NME3 gloeobacter
21	262	6.5	1270	Q6UYJ9	Q6UYJ9 burkholderi
22	238	5.9	2468	Q9I2M3	Q9I2M3 pseudomonas
23	226.5	5.6	2082	Q74DR4	Q74DR4 geobacter s
24	214	5.3	1800	Q9L948	Q9L948 pseudomonas
25	214	5.3	8682	Q8BRG2	Q8BRG2 pseudomonas
26	205.5	5.1	2888	Q6J0T3	Q6J0T3 burkholderi
27	205	5.1	2215	Q7WBN0	Q7WBN0 bordetella
28	202.5	5.0	2174	Q92UU8	Q92UU8 rhizobium m
29	198.5	4.9	3346	Q7WN54	Q7WN54 bordetella
30	197	4.9	2016	Q8TJS9	Q8TJS9 methanosarc
31	196	4.9	494	Q63J05	Q63J05 burkholderi

32	189	4.7	3972	2	P73139	P73139 synechocyst
33	186.5	4.6	3229	2	Q63UE4	Q63UE4 burkholderi
34	186.5	4.6	4199	2	P74440	P74440 synechocyst
35	186	4.6	905	2	Q7NCI3	Q7NCI3 gloeobacter
36	185.5	4.6	818	2	Q9NLP0	Q9NLP0 bos taurus
37	185.5	4.6	997	2	Q9Z4I1	Q9Z4I1 bacillus sp
38	184.5	4.6	998	2	Q9HLQ9	Q9HLQ9 thermoplasma
39	184	4.6	805	2	Q88GZ6	Q88GZ6 pseudomonas
40	183	4.5	2523	2	Q8TJS8	Q8TJS8 methanosarc
41	182.5	4.5	848	2	Q9RK65	Q9RK65 streptomyc
42	182	4.5	1070	2	Q8XW00	Q8XW00 ralstonia s
43	182	4.5	1904	2	Q83119	Q83119 lactococcus
44	181.5	4.5	428	2	Q6ZE71	Q6ZE71 burkholderi
45	181.5	4.5	1660	2	Q6MK63	Q6MK63 bdellovibri

ALIGNMENTS

RESULT 1

ID	Q82M04	PRELIMINARY;	PRT;	882 AA.
AC	Q82M04;			
DT	01-JUN-2003 (Tremblrel. 24, Created)			
DT	01-JUN-2003 (Tremblrel. 24, Last sequence update)			
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)			
DE	Putative endo-1,4-beta-glucanase.			
GN	Name=celA3; OrderedLocNames=SAV1856;			
OS	Streptomyces avermitilis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomyces; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=33903;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NA-4680;			
RX	MEDLINE=22608306; PubMed=12692562;			
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,			
RA	Sakaki Y., Hattori M., Omura S.;			
RT	"Complete genome sequence and comparative analysis of the industrial			
RT	microorganism Streptomyces avermitilis."			
RL	Nat. Biotechnol. 21:526-531(2003).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;			
RX	MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;			
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,			
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,			
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;			
RT	"Genome sequence of an industrial microorganism Streptomyces			
RT	avermitilis: deducing the ability of producing secondary			
RT	metabolites."			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).			
DR	EMBL; AF005028; BAC69567.1; -.			
DR	HSSP; P07986; 1EXG.			
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR001919; Bac_celose-bind.			
DR	InterPro; IPR008965; Cellul bind.			
DR	InterPro; IPR002860; Glyco Hydro_BNR.			
DR	InterPro; IPR011040; Sialidase.			
DR	Pfam; PF02012; BNR; 10.			
DR	Pfam; PF00553; CBM_2; 1.			
DR	SMART; SM00637; CBD-II; 1.			
KW	Complete proteome.			
SQ	SEQUENCE 882 AA; 92152 MW; A3C8E753CE69F13D CRC64;			

Query Match 60.2%; Score 2429.5; DB 2; Length 882;

Best Local Similarity 59.8%; Pred.No. 2.1e-130;

Matches 438; Conservative 111; Mismatches 174; Indels 19; Gaps 8;

QY 6 YTSNVAIGGGGFDVFNNEGAPGILYVRTDIGMYRDAANGRWITPLLDVGVGNMGY 65

Db 39 YSKNARVDGGFPGVFNRSKKNLAYARTDIGGAYRMAESSKNTWTFLLDSVGVGDWGH 98

718 RRVYGTNGRGIVYGD1-----GGAPSG 740
 749 GRVYATNGRGVYGD1TSDTGGTDPG 775

Db

RESULT 3

Q9A0H0 PRELIMINARY; PRT; 996 AA.

ID Q9A0H0
 AC Q9A0H0
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Glycosyl hydrolase 5 (Fragment).
 OS Caldicellulosiruptor sp. Tok7B.1.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
 OC Caldicellulosiruptor.
 NCBI_TaxID=80339;
 RN [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=201711169; PubMed=10706665;
 RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
 RG Bergquist P.L.;
 RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
 thermophile Caldicellulosiruptor isolate Tok7B.1.";
 RL Curr. Microbiol. 40:333-340(2000).
 EMBL; AF078038; AAK06388.1; -;
 DR HSP; Q06851; INBC.
 DR GO; GO:0016787; F.hydrolase activity; IEA.
 DR GO; GO:0005975; P.carbohydrate metabolism; IEA.
 DR InterPro; IPR001956; CBD_3.
 DR InterPro; IPR008965; Cellul_bind.
 DR InterPro; IPR002860; Glyco_hydro_BNR.
 DR InterPro; IPR011040; Sialidase.
 DR Pfam; PF02012; BNR; 9.
 DR Pfam; PF00942; CBW_3; 1.
 DR ProDom; PD001947; CBD_3; 1.
 DR Hydrolase.
 KW Hydrolase.
 FT NON TER
 SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED2F3C614 CRC64;

Query Match 51.1%; Score 2062.5; DB 2; Length 996;
 Best Local Similarity 52.6%; Pred. No. 2.1e-109;
 Matches 387; Conservative 96; Mismatches 238; Indels 15; Gaps 9;

QY 4 QPVTWSNVAI-GGGGFVDGVFNEGAPGILYVTRTDIGGMVYRMDAANGRIPLLDWVGNN 62
 DB 35 QPVTWSNVAI-GGGGFVDGVFNEGAPGILYVTRTDIGGMVYRMDAANGRIPLLDWVGNN 94
 QY 63 WGYNGVVSIAADPINTNKV-WAAGVMYTNWDNDGAILRSSDQATWITLPPFKLGN 121
 DB 95 WNLGVSATDPDVPKXVILACQGYTNSWTDMDGAILRSTDEGTFEITLPPFKLGN 154
 QY 122 MPRGNGERLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 181
 DB 155 MPRXNLGERLAIDPNNRILYLTREGXGLWKSSEDYGVSWKKVTSFPNPGTYIEDNCPN 214
 QY 182 GYQSDIQGVVWAFDKSSSLGQAKTIIFGVADPNPVFWSRDGGATQAVPGAPTGI 241
 DB 215 DYLNHITGVVWVFDPTSGRPGGXKILYVGVADKTTSYIYTKDGGQWALPEQPTGLL 274
 QY 242 PHKGVEPDVNVHLYIATNTGGYDGGSDGVWKFVSTSGTWTRISVPVSTDTANDYFGYS 301
 DB 275 PQRKLSS-DGMLIYTSNTGYPNGDYGEVRYNTKTGEWKNISPMAAQDT---YFGYG 330
 QY 302 GLTIDQHPNTIWAQISWPTTIIPRSTGGATWTRIWDWTSYNRSLRYVLDISAEP 361
 DB 331 GLAVDAQNPVKVWVAALSSWPDYIWRSDGGETWKCIEWMNGYNRNLTHYNMDISAAP 390
 QY 362 WLTFG-VQNPVPVPSKLGWDEMAIADPNSDRMLYGTGATLYATNDLTKWDGSGQIHI 420
 DB 391 WLMFGXTBTPPEVPSKLGWVGWGLEIDPNPSDMXLYGTGAXLYGCDLLTKWDGQNIIT 450
 QY 421 APVVKGLEETAVNDLISPPSGAPLISALGDLGGTFHADVTAVPSTTIFTSVPVTTGTSVDY 480


```

QY 359 ABPWLTFQVQNPVPSP---KLGWDEAMADFPNSDRMLYGTGATLYATNDLTWKDWSG 415
D 379 LAPW---IGNIVDVTPGNLQICWMMESLSIDFFDSNHNHLYGTGTEIYSGRDLWKDWSA 434
QY 416 GQIHIAPMVKLESTAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTG 475
D 435 HNVTIKSLADGVEETSQALISPPSGPLVSAIGDVGVGHQSLTRAPSAQFTNPKWSTS 494
QY 476 TSVDYAEINPSIIIVRAGSDFPSSQPNDRHVAFTSDGKNWFGQSGEPGVTGGTVAAASD 535
D 495 ADIDFAGNPNPNVRIIGTGDS---GKQVAISSDYGVTVNQHQGAPDNVQGGKVAISAD 551
QY 536 GSREVPWAPDGPQVAVYAVGFSGNSWAASQGVPAANAQIRSDRVNPKFTFYALNGTFFYRSTD 595
D 552 ADIILWRTNGNG-----VMVSRNQATFNI FYGASGKTFYVSTD 589
QY 596 GGVTFQPVAAAGLPSGGAVGVMFHVPKGEGLWLAASSGLYHSTNGSGSWAISTGVSSAV 655
D 590 NGKTFSAHSGLSGATSVDITVH--PSVSGDIWASTDKGLFHSITDSGATFSAISGITOAW 647
QY 656 NVGFGKSPAGSSYPAVVGTGIGVGTGAYRSDDCGTTWVLINDDOHQYG-NWQQAITGDH 714
D 648 GVALGAPRSTGGYPAVFAAANYGVEAYFRSDDRGVNWKINDAAHGFGAASANCMAADP 707
QY 715 ANLRVYVGTNGRGIVYGDIGGA 737
D 708 RVYGRVYVGTNGRGIFYGDVAGS 730

RESULT 8
O74170
ID O74170 PRELIMINARY; PRT; 856 AA.
AC O74170;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Avicelase III.
GN Name=avicii;
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai M., Takada G., Kawaguchi T., Sumitani J.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB015511; BAA29031.1; -.
DR PIR; T00349; T00349.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR002860; Glyco hydro_BNR.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 7.
DR Pfam; PF00734; CBM 1; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
SQ SEQUENCE 856 AA; 89820 MW; BE085983AF60ED76 CRC64;

```

```

Query Match 41.6%; Score 1680; DB 2; Length 856;
Best Local Similarity 46.2%; Pred. No. 1.2e-87;
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 1 ATTPQYTSNVAI--GGGFGVDGIVFNEGAPGLYVRTDIGNMYRWDAAANGWIPLLDWVG 59
D 21 AASQAYTKNVVTVGGGGFTGIVNPSAKGVAVARTDIGAYRLN--SDDTWTPPLMDWVG 79
QY 60 ---NNWNGYGVGSAADPINTNKVAAVGYMTNSWDPNDGAILRSSDQGTWQTPLPF 116
D 80 NDTWHD--GIDALATDPDVTDRVYVAVGYMTNEDWPNVSGILRSTDQGDWTETKLPF 136

```

```

QY 117 KLGNNMPCRGMERLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWQMTNFPDVGYYIAN 176
D 137 KVGGNMPCRGMERLAVDPNKNISIIYFGARSCHGLWKSTDYDGTWNSVTSFTWTGYFQD 196
QY 177 PTDITCYQSDTQCVVWVAFDSSSSSLGQASKTIIFYGVADPNPNPVFVSRDGGATWQAVPGA 236
D 197 SSST--YTSDPVGIAWVTFDSTSGSGSATPRIFYGVADAGKSVFKSESDAGATWAWVSGE 254
QY 237 PT-GFPHKGVDPVNVHVIYATSNTPGYPDGGSSGDMVKFSVTSQVWTRISPVSPSTDAN 295
D 255 POYGFPHKGVLSPEKTLIYISYANGAGPYDGTNGTVHKYNTISGVWTDISP---TSLAS 311
QY 296 DYFGSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGGATWTRINDWTSYPNRSRYVL 355
D 312 TTYVGGLSVDLQVPGTLMVAALNCWMPDELIIFRSTDSGATWSPIEWENGYPSINYYSY 371
QY 356 DISAPFWITFGVQPNP--PVPSFKLGWDEAMADFPNSDRMLYGTGATLYATNDLTWKDS 414
D 372 DISAPWITQDITSTDDQFPV---RVGMVVEALADFPDSNHNHLYGTGLTVYGGHDLTNWDS 428
QY 415 GQIHIAPMVKLESTAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTT 474
D 429 KINVTVKSLAVGIEWAVLGLITPPGFPALLSAVGDDGGFYHSDLDRAFPNQATHTPYGT 488
QY 475 GTSVDYAEINPSIIIVRAGSDFPSSQPNDRHVAFTSDGKNWFGQSGEPGVTTCGTVAAASA 534
D 489 TNGIDYAGNKPNSIVRSGASD--DYPT---LALSSNFGSTWYADYAASTSTGTGAVALSA 543
QY 535 DGSRFVWAPDGPQVAVYAVGFSGNSWAASQGVPAANAQIRSDRVNPKFTFYALSNGTFFRST 594
D 544 DGDITVLLMSSTSGALVSKSQG---TLTAVSSLPFGAVIASDKSDNTVYFGSGAGAIYVSK 600
QY 595 DGGVTFQPVAAAGLPSGGAVGVMFHVPKGEGLWLAASSGLYHSTNGSGSSWSAI--TGVSS 653
D 601 NTATSTFTKTVS--LGSSTTVNAI--RAHPSIAGDVWASTDKGLHSTDYGTFTQIGSVYA 658
QY 654 AVNVGFGKSPAGSSYPAVVGTGIGVGTGAYRSDDCGTTWVLINDDOHQYGNWQQA--IT 711
D 659 GWSFGFGKASSTGTVVYVIGPFTIDGAAGLFPKSEDAGTWQVTSIDASHGFGS--GSANV 717
QY 712 GDHANLRVYVGTNGRGIVYGDIGAPSG 740
D 718 GDLQTYGRVFRGHERPCHLLRSQREPAG 746

RESULT 9
Q729M8
ID Q729M8 PRELIMINARY; PRT; 838 AA.
AC Q729M8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cel74a.
GN Name=cel74a;
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QM6a;
RA Foreman P.K., Brown D., Dankmeyer L., Dean R., Diener S.,
RA Dunn-Coleman N.S., Goedegebuur F., Houfek T.D., England G.J.,
RA Kelley A.S., Meerman H.J., Mitchell T., Mitchinson C., Olivares H.A.,
RA Teunissen P.J.M., Yao J., Ward M.;
RT "Transcriptional Regulation of Biomass-Degrading Enzymes in the
RT Filamentous Fungus Trichoderma reesei.";
RL J. Biol. Chem. 0:0-0(2003).
DR EMBL; AY281371; AAP57752.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000254; CBD_fungal.

```


Qy	423	MVK-----GLEETAVNDLISPPGAPLIISAIGDLGGFTHADVAVPSTIFTSVPVTTGT	476
Dd	385	PVKMIFENRGLLEETVVLQVLVPIGERPLLSAIAIDWGGFRHESLDTPSSMY-KPLKWTSL	443
Qy	477	SYDYAEALNPSIIVRAGSFDPPSQPNDRHVAFSTDGGKNWFQ-CSEPGGVTTGG--TVAAS	533
Dd	444	GIAFAIQNSKFVARVHTY---TYP---FLSYSDEGGINWREIETVPEGITDGGRLSLAYS	497
Qy	534	ADGSRFVNAPGDGPQPVVAVFGNSWAASQGVA---NAQIRSDRVNPKTVAL---SNG	588
Dd	498	NDGKTLVSWSPAN--HEVIVSSDKGSKMKALSVPEPFNFPAADPVPNPSKFYIPDNKG	555
Qy	589	TFYRSTDGGVTFPQVAAGLPSSGAVGVNFHA----VPGKEGDILWLAAAS-SGLYHSNTGGS	643
Dd	556	DFLISEKDGGKSPMK-GAKLPFDNNWSLYSFVPVLAPDREGDIWLALQNWNLRSXDGGI	614
Qy	644	SMSAITGYSSAVNVGFSGKSAPCSSYPVFFVVGITGGVTGAYRSDDCGTTWLINDPDHQY	703
Dd	615	TEERLGNDVIADVIGFGAPKPTDPYAIYLNGMWNVGYIFMSTDEGKTWMRINNDDKHQF	674
Qy	704	GNWGQAITGDHANLARVVINGNRGIVYGDI	734
Dd	675	G-WHYMTGDMMNEFGRIEFLGTGEGRGIIIVGEV	704

RESULT	ID	QBP9U5	PRELIMINARY;	PRT;	751 AA.
AC	QBP9U5;				
DT	01-OCT-2002	(T+EMBLrel. 22, Created)			
DT	01-OCT-2002	(T+EMBLrel. 22, Last sequence update)			
DT	01-MAR-2004	(T+EMBLrel. 26, Last annotation update)			
GN	Cellulase.				
GN	OrderedLocusNames=XCC1752;				
OS	Xanthomonas campestris (pv. campestris).				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;				
OC	Xanthomonadaceae; Xanthomonas.				
OX	NCBI_TaxID=340;				
RN	[1]				
RP	SEQUENCE FROM N. A.				
RC	STRAIN=ATCC 33913 / NCEPB 528;				
RX	MEDLINE=22022145; PubMed=12084217; DOI=10.1038/417459a;				
RA	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,				
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,				
RA	Alves L.M.C., do Amaral A.L., Bertolini M.C., Camargo L.E.A.,				
RA	Camarote G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,				
RA	Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,				
RA	Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,				
RA	Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,				
RA	Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,				
RA	Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,				
RA	Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,				
RA	Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,				
RA	Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,				
RA	Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,				
RA	Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,				
RA	Setubal J.C., Kitajima J.P.;				
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing				
RT	host specificities.;"				
RL	Nature 417:459-463(2002).				
DR	EMBL; AE012276; AAM41043.1; -;				
DR	InterPro; IPR002860; Glyco hydro BNR.				
DR	InterPro; IPR000169; Pept Cys acSite.				
DR	InterPro; IPR011040; Sialidase.				
DR	Pfam; PF02012; BNR; 7.				
DR	PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.				
KW	Complete proteome.				
SQ	SEQUENCE 751 AA; 81140 MW; 7378B6AA0B5698A4 CRC64;				

Query Match

Best Local Similarity 35.0%; Score 1145; DB 2; Length 751;

Matches 265; Conservative 121; Mismatches 295; Indels 78; Gaps 21;

Qy	1	ATTQPTYSNVAICGGGFVDGIVNEGAPGILLYVRDITGGMYRWDAANGRWITPLLDWVGH	60
Db	41	ATSGPYQWRSVAICGGGFVTGVLFHPAERGLAYARTDVGGAYRWDAQOQWALTDLWGA	100
Qy	61	NNMGYNCGVSI TAADPI NTNKVMAAVGMTNSWDNDGAILRSSDCGATWQITPLPPKLG	120
Db	101	DDNLMIGDIFAVDPADADALYLAAGYVWHERAGY-AAVLSFNGRIFERADLFPKLG	159
Qy	121	NMPGRGMRGLAVDPNNNDNLIYPGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT	180
Db	160	NQLGRANGERLARVDPHGRVLLLSRDA-GLWRSDDRGAHWAKVASFPPDAALAGATARNH	218
Qy	181	TGQSDTQGVVWVAFDKSSSLGQASKTI FVGVDPPNPFWRSRGGCATWQVPGAPTGF	240
Db	219	VGREQAV-GIAFVVFDAAAGNTGPTPTRIYVGVSTEQSLYVSEDAGRSWAPVAGQPRGL	277
Qy	241	IPHKGVDPVNVHLYIATSNITGGPYDGGSSGDVWKFVSTSGTWTRISPPVSTDTANDYFY	300
Db	278	RPSHMAGSGDH-WYLSYGQDPGLMAGGALKWETPAGRWREISPIPO-PASGDGFGW	335
Qy	301	SGITIDRQHPNTIMVATOISWMPDTI IFRSTDGGATWTRIWDWTSPYKNSLYVLDISAE	360
Db	336	GAVAVDPQOPQVILLASTFRRTPRDELYRSVDGKGHWAPL-----LADAVFDHSA	386
Qy	361	PWLTFFGVQPNPPVPSPKLWMDMAWIDPNSDRMLYGTGATLYATNDLTKWDSGQIHI	420
Db	387	PW-TAHATPH-----WNG-ALAIIDPFGNHALFVTGYIWSARNJ-----QDFA	428
Qy	421	AP-----MVKLEETAVNDLISPPSGAPLISALGDLGFGTHADVTAVPSTIFTSPVF	472
Db	429	APQRLQWQDQGLEETVPLDLLSPWAGAHLLSALGDI DGFPHDELRA-QLYVAGPRL	487
Qy	473	TTGTSVDYAEINPISIIVRAGSFDSSSQPNDRHVPASTDGGKWFO-GSPGVTTCGTVA	531
Db	488	TNGESIDAAGQAPQWVYRSGTV-RDRRNEIRALYSRDKGKQMTAFASPPAGQAGASIA	546
Qy	532	ASADGSRFVWAPGDPGPVVYAVGFGNSWAASQGVPAQAQIRSDRVPNPKTFYA--LSNGT	589
Db	547	IGADAAQVWVAPERGGN--WRTSDFGAQQRVDGLPDTAVVMDRVDARRTAVDVASGQ	604
Qy	590	FYRSTDGGVTFQPVAAGL-----PSSGAVGVMFHAVPGEGDLMLAASS	633
Db	605	LYESTDAARSFR--ATGVQVGSPPARDERTRPQLRPDPWRAGVVYLASPGK-----	652
Qy	634	GLYHSTGGSSWSAITGVSSAVNTVFGKSAFGSSYPVAVFVGTIGGVTGAYRSDDCGTTW	693
Db	653	GVNRWQDG--TLQVLSQDPDEARSLIGKALRAGAPPALLYLAGRVQGVDFRSDGGVOW	710
Qy	694	VLINDDOHGVNWCQAITGDHANLRRVYIGTNGRIGVYGD	733
Db	711	QRINDDAHRFRG-PYSTGDPRIAGRVVPATGGRGIFYGD	749
RESULT 15			
Q8J0D2	Q8J0D2	PRELIMINARY;	
ID	Q8J0D2	PRT;	812 AA.
AC	Q8J0D2		
DT	01-MAR-2003	(TrEMBLrel. 23, Created)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)	
DT	01-JUN-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Oligoxylolog can reducing end-specific cellobiohydrolase.		
OS	Geotrichum sp. M128.		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae;		
OC	Geotrichum.		
OX	NCBI_TaxID=203496;		
ON	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=M128;		
RX	MEDLINE=22359052; PubMed=12374797; DOI=10.1074/jbc.M208443200;		
RA	Yaoi K., Mitsuishi Y.;		
RT	"Purification, characterization, cloning, and expression of a novel		

Query Match	28.4%	Score 1145;	DB 2;	Length 751;	KC
Best Local Similarity	35.0%;	Pred. No. 3.4e-57;			RX
Matches 266;	Conservative 121;	Mismatches 295;	Indels 78;	Gaps 21;	RA
					RT
SRR191NW226; MEDLINE=22359052; PubMed=12374797; DOI=10.1074/jbc.M208443200; Yaoi K., Mitsuishi Y.; "Purification, characterization, cloning, and expression of a novel "					RT

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:49:28 ; Search time 28.4312 Seconds
(without alignments)
1942.949 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATQPYTWSNVAIGGGFVD.....YIGTNGRIVYDGGAPSG 740

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgm2_6/ptodata/1/iaa/5A-COMB.pep.*
2: /cgm2_6/ptodata/1/iaa/5B-COMB.pep.*
3: /cgm2_6/ptodata/1/iaa/6A-COMB.pep.*
4: /cgm2_6/ptodata/1/iaa/6B-COMB.pep.*
5: /cgm2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
6: /cgm2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219.5	5.4	2736	4	US-09-252-991A-30227
2	181.5	4.5	3892	4	US-09-328-352-5503
3	167.5	4.2	1164	4	US-09-902-540-12627
4	166.5	4.1	1300	4	US-09-902-540-9932
5	162.5	4.0	772	2	US-08-410-784A-5
6	159.5	4.0	1060	4	US-09-248-796A-14123
7	158.5	3.9	774	3	US-09-346-237-8
8	157.5	3.9	688	4	US-09-489-039A-9813
9	154	3.8	1548	4	US-09-252-991A-22301
10	152.5	3.8	1751	3	US-09-136-574A-44
11	152	3.8	3472	4	US-09-408-020-4
12	151.5	3.8	776	3	US-09-346-237-4
13	148.5	3.7	750	6	5457037-3
14	148.5	3.7	750	6	5457037-3
15	148.5	3.7	751	6	5457037-5
16	148.5	3.7	751	6	5457037-5
17	148.5	3.7	776	3	US-09-346-237-7
18	147	3.6	3623	4	US-09-341-461-2
19	146	3.6	1187	4	US-09-949-016-6513
20	146	3.6	1749	4	US-09-640-419C-28
21	146	3.6	2169	4	US-09-949-016-6930
22	143	3.5	894	3	US-08-362-525-22
23	143	3.5	894	3	US-08-971-692-15
24	143	3.5	4861	4	US-09-919-497-70
25	142	3.5	1216	4	US-09-134-000C-5130
26	141	3.5	720	3	US-09-296-284-25
27	141	3.5	754	3	US-09-296-284-4
					Sequence 30227, A
					Sequence 5503, Ap
					Sequence 12627, A
					Sequence 9932, Ap
					Sequence 5, Appl
					Sequence 14123, A
					Sequence 8, Appl
					Sequence 9813, Ap
					Sequence 22301, A
					Sequence 44, Appl
					Sequence 4, Appl
					Sequence 4, Appl
					Patent No. 5457037
					Patent No. 5457037
					Patent No. 5457037
					Sequence 7, Appl
					Sequence 2, Appl
					Sequence 5513, Ap
					Sequence 28, Appl
					Sequence 6930, Ap
					Sequence 22, Appl
					Sequence 15, Appl
					Sequence 70, Appl
					Sequence 5130, Ap
					Sequence 25, Appl
					Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-252-991A-30227

; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30227
; LENGTH: 2736
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match 5.4%; Score 219.5; DB 4; Length 2736;

Best Local Similarity 21.2%; Pred. No. 6.9e-07;
Matches 190; Conservative 94; Mismatches 324; Indels 289; Gaps 48;

QY	59	GNNWGY	-----NGV--SIAADPI-NTNKVWA-----AVGMYTNSWDNDGAILR-SS	103
DB	1305	SGSNWTFTPATPLANGTVVNAVAQDPAGNTPGQGSTTVDVAVPTPVNFSNGLLNGTA	1364	
QY	104	DCGATWQITPLPKLGGNMPGRGMGERLAVDPNNNDILYFCAPSGKGLWRSTDGATWSQ	163	
DB	1365	EPGSTVLT-----DCN--GNPIGQTTA-----DGSNWSFTPG-----SQ	1398	
QY	164	MTNFPDVGTYI-ANPTDTTGYQSDIQGVVWVAFDKSSSLGQ---ASKTIFVGVDNPNP	219	
DB	1399	LPN-----GTVMVNTASDAAGNTS---APATTTVDSSLPSIPQVDPSPNGSVISGTADAGNT	1451	
QY	220	VF-----WSRDGGATWQAVPCAP-----	242	
DB	1452	IIITDGNPNIGQVTDAGSGNWSFTPGIPLPDGTVMNVARSNPNVDSAPAVITVDGVAP	1511	
QY	243	HKGVFPVNVHLVIATNTGGPY---DGSGDVWKFVST-SGTWTRISPPV-----	289	
DB	1512	AAPVIDPSNGTEISGTAEAGATVILTDGGNPGIQATADSGNWTFTPGIPLANGTVNA	1571	
QY	290	-STDTANDYFGYSLTIDRHPNTIMVATQISNWPDTIIFRSTDGGATWTRIDWTSVFN	348	
DB	1572	VAQDPAGNTSGPASVTVDAIAP---PAPVINPSNGVVISGTAEAGAT-VILTDGNGNP-	1625	


```
QY 349 RSLRYVLDISAEPMLTFTGVQPNPPV-----PPSKLGMWDE-----AMAI 387
Db 1626 -----IGQVTDAGSGKWAFPTATPLANGTIVINALQAAGNNSPTSATVDSLAPAAPI 1680
QY 388 DPFNSDRMLYGT---GATLYATNDLTWKDSSGQIHIAPMVKGLEBETAVNDLISPPSGAPL 444
Db 1681 DPNSSG-VIAGTAEGAGATVILT-----DNGN-----PIQVTDAGSGNWSFTT-GRP 1727
QY 445 -----ISALG-DLGGFTHADV-TAVPSTIFTPSVF---TTGTSV-DYAEALNPSIIVRAGSF 494
Db 1728 SNGTVVNAQAAGNTSGPASTTVDVAPAPVDPDPSNGSVIAGTAEGATVILTDGG 1787
QY 495 DPSSQPNDRHFAFSTDGKNN-----FQSBFGVTTGGT----- 529
Db 1788 NPIGA-----TADSGNSFTPGTPLANGTVINAQAQDPAGNTSGPTSTTVDAVAPA 1840
QY 530 -----VAASADSGSRFWAGDPGQPV---VYVGFNGSNAASGCV-ANQAIRS- 574
Db 1841 TPVVPNSGSVIAGTAEGATVILTDGGNPIGQVTDAGSGN-WSFTPGTFLANGSVINA 1899
QY 575 -----DRNPKT-FYALSNGTFYRS-----TDGGVTFQPVAAAG 606
Db 1900 LAQDAAGNTSGPASTTVDVAPATPVLDPNSGTVISGTAEGATVILTDGG--GNPIRQA 1957
QY 607 LPSSGAVGVMFHAPVKEGDLWLAASSGLYHSTNG--GSSWSAITGVSSAVNVGFGKAP 664
Db 1958 TADSGNSWFTPGTPLTNGTVINAQAAGNTSGPVSTTVDAVAPATPVIDPSNGVKLS 2017
QY 665 GSSYPAPVFWGTIGG-----VTGAYRSDCGTT----- 692
Db 2018 GTAEPGVRVILTDGNGNPIGOTLADSGNWTFTPGTPLANGTVVNAQAQDPAGNTSGPAS 2077
QY 693 -----WVLINDQHQYGNWQAITGDHANLRVVGINGRIGVYDGGAPSG 740
Db 2078 TTVDTVAPATPVINASN-----GSVITGT-AEVGAKVILTDGNGNPIGETTADGSG 2127

RESULT 2
US-09-328-352-5503
; Sequence 5503, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5503
; LENGTH: 3892
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5503

Query Match 4.5%; Score 181.5; DB 4; Length 3892;
Best Local Similarity 21.5%; Pred. No. 0.00067;
Matches 204; Conservative 90; Mismatches 344; Indels 309; Gaps 54;

QY 7 TWSNVAIGGGGFVDG-----IVFNEGAPGILYVR-TDIGMYRWDAANGRWIPLLDW 57
Db 543 TWT-VSVPGSLVADADTKIDAKVFTTDAAGNSSTVNDTQI---YTLDTA-APAAPVIDP 597
QY 58 VGNWNWYNGV---VSIADAPINTNKVAAVGMYN-----SND-PNDGAILRS 102
Db 598 V-----NGTPITGTAEPSGTIV-----TYPNGDTATVAGPDGWSVNPNG--LND 643
QY 103 SDQATWQITPLPFKLGN--MPGGMGERLAVDNN-----DNILYFAPSG 148
Db 644 GDEVEAIATDP-----AGNPSLPGTATVD--AVGNTDGVNFTVSDVADNVINASEASG 696
QY 149 K-----GLWRSTDSGATWSQMTNFPDVGTVIANPTDTTGYSQDIQGVVWVAFDKSSSLGQ 204
```

```
Db 697 NVTVTGLKNVPADAANTVTVVINGQTYTATVDSTAG-----TWTVSVPGSOLITAD 748
QY 205 ASKTI-----FVGADPNPNPVFW 222
Db 749 ADKTIADKVFTDAAGNSSVNDTQTYTITDTPADAPVNPVNGTDPITGTAEPSGTIV 808
QY 223 SRDGATWQAVPGAP-TGFI PHKGVDPVNVHLYIATSNWGPYDGSNGVWKFVSVTSGT 281
Db 809 TYPDGSITTVVAGPDGTWTPNPNGLNDG-DKVTATATDPAGNP-----SLPGT 855
QY 282 WTRISPVSTDTANDYFGYSGLTIDR-----QHPNTIMVATQI 319
Db 856 ATVDVAGNTDGVN--FTVDSVTADNVINASEASGNVTVTVGLKNVPADAANTV-VTVVI 912
QY 320 SWMPDTIIFRSDGGATWRIWDTSYPNRSIRYVLDISAEPLWTF----- 365
Db 913 NGQTYTATVDSTAG--TWT-----VSVFGSLDTADADTKIDAKVFTTDAAGNSSVNDTH 965
QY 366 -----GVQPNPVPSPKLGWMDMAIDPFNSD-----RMLY--GTGATLYATND--- 408
Db 966 TVTVDTVAPNAPV-----LDPINATDPVSGAEPGSTVTVTPDGTATVAVGPDGSW 1018
QY 409 -----LTKWDSGQIHI-----APMYKGLEETAAND-----L 435
Db 1019 SVPNPGLNVGDVTATATATDPAAGNTSLPCTGTVSADITAPVV-ALDDVLTNDSTPALGT 1077
QY 436 ISPPSGAPLISA-----LGDLGGFTHAD-----VTAVPSTIFTPSVFTTG----- 475
Db 1078 VNDPTATVVNVVDGYDPAVNGD-GTWTLANDTLPALTDGPHITVATDAAGNAGTDT 1136
QY 476 --TSVDYAEALNPSIIVRAGSFPDPSQPNDRHFAVSTDGKWNWFGSGEPGVTGGTVAAS 533
Db 1137 AVVTIDTTAPNAPVLDPINATDP-----VSGTAEGST-VTVTPDGT--ATVWAG 1185
QY 534 ADGSRFWAGPDGPQPV-----VYVGFNGSNAASGCVPANAIQIRSRVNPKEFYALSNGT 589
Db 1186 TDGS---WSVPNPGNLVDGTVTAT--ATDPAGNTSLPCTGTVSADITAP--VVALDDVL 1238
QY 590 FVRSYTDG--GVTFQFVAAGLPSGAGVGVMFHAPVKEGDLWLAASSGLYHSTNGSGSWA 647
Db 1239 TNDSTPALGTVNDPTATVVVN--VDGTYDPAVNGDGTG-TWTLANDTLPVLADGPHIT-IT 1294
QY 648 ITGVSSAVNVG-----FGKSAPGSSYPAVFVGTIGGTVGAYRSDDC-----GTTW 693
Db 1295 VTATDAAGNAGTDTAVVITDITAPNA--FVLDPINATDPVSGTAEGSTVTVTPDGTTA 1352
QY 694 VLI--NDDQHQYGNWQAITGDHANLRVVIQTN-----GRGIVYGD 734
Db 1353 TVVAGTDGWSVNPENLVDGDTVTATATDPAAGNTSLPCTGTVSADI 1399

RESULT 3
US-09-902-540-12627
; Sequence 12627, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 12627
; LENGTH: 1164
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-12627
```


Query Match	4.2%	Score 167.5	DB 4	Length 1164
Best Local Similarity	19.8%	Pred. No. 0.0013		
Matches	173	Conservative 89	Mismatches 333	Indels 279
Gaps	42			
Qy	6	YTRSNVAIGGGGFVDGIFVNEGAPGILYVRTDIGMYRWDAANGRWIPLG--- <td>63</td> <td></td>	63	
Db	179	WAMGRNAAGQLG--DGTTLDRARPVAVPGLTSTVAV-----AAGSHVLALRSQGSVM-AW	231	
Qy	64	GYN-----GWYSIAADPINTKNKVAAGVMYTN-----WDPNDGAILRSSD	104	
Db	232	GYNALGQLDGTTVDRLTTPRVSLGAVAVAAAGSYFSMALQSDGTVMWCEGFGQLGD	291	
Qy	105	QGATWQITLPIFKLGNMPCRGMERLAVDPNNILYFGAPSKGLWRSTDGATWSQM	164	
Db	292	GGGVORLSPVRVEGLANITRVAAGSAHALAVRGDGTVMWADNGEGQL-----GDGSWADR	347	
Qy	165	TNFPDVGTYIANPDITTYGQSDIQGVVWAFDKSSSLGQASKTI-----FVGADPNNPV	220	
Db	348	-----FRPVQVFG-----IQGITAVSGGRSHSMALQDGTFRWVINGYQQLGDGT	393	
Qy	221	FWSRD-----GGATWQAVPQAP	237	
Db	394	LTSVRPALLPGLTGIQALHPSHLHLALHADGTLRGWYNRFQGLGLGAAGSAVPQV	453	
Qy	238	TGFIPIKHGVDPVN-----HVLVIATSN-----GGPYDGSSGDWVKFSVTSQWT--RISP-	287	
Db	454	RGI-----GRVDRLSVGRAHTLMVRADGTVMWANGSGQLGD-----GTSTHRTGVP	501	
Qy	288	-----VPSTDYANDYFGYS-----GLTIDRQHPNTI-----MVA	316	
Db	502	SVPGVPCVRSVASGIQHSALACDGTVMWAGANARQQLGDGTTTPRTVPLVEGLRGVVA	561	
Qy	317	TOISWMPDTIIFRGTGCGATWTRITWDTSYPNRSLRYVLDISABEPLTFGVQPNPPVSP	376	
Db	562	--LAAGDASVALRADGS-----VMSWGG-----NASQLGDGTLTDRALPA-	601	
Qy	377	KLGMWDEMAIDPFPNSDRMLYGTGATLYATNDLTKW-----DSGQIHIAPM-VKG	426	
Db	602	HVEHLSNAVPVPVGETHALAVGEDALWA-----WGANGSQQLDGGGTAPSLTFVRVKG	655	
Qy	427	LEETAVNDLISPPSGAPLISALGDLGGTHADVTAVPTIF---TSPVFTTCTSYDYAEL	483	
Db	656	LERVA-----SVAAGRAFSVAVRDDG-----TAMWAGTNPSQLDGTGNTNHAKS	698	
Qy	484	NPSITVRAGSDPSPQPNDRHVAFSTDCGKNWFOG-----SPPGG-----	523	
Db	699	VPNQVTSLKGTRAMSAHAHVVALSAD-GTVMWGDNTLQQLDGSSTPGWMPQVPEL	757	
Qy	524	-----VTTGG---TVAASADGSRFVWAGDPGPVVYVAVGFNGSMAASQGVPAANAQRSDR	576	
Db	758	RGVEAVAAGEQFTVAVLDDGTARAGSNEYGO-----LGDGKTGPQLTPMAVKIEGPK	810	
Qy	577	VNP--KTFYALNNGTFYRSTGGV-----TFQVAAAGLPSSGAGVWFHAPVPGEGDLW	628	
Db	811	IRPPVTRVWAGQHAVVQMSDGTQVTWGDNTFQQLGDGTTTTRAVPLTVIGLASV-----	865	
Qy	629	LAASSGLYHS---TNGSGSWAITSVCSSAVNVVFGKSPAGSSYPV-----VFVY	674	
Db	866	VAVSSGAHSLALLSDGTVWA--WGAN-----GFQLGDGTSTPTRPRPVQVAGLDVVAI	918	
Qy	675	GTIGVTCAYRSDDCGTTTWLINDDOHQYGNWQ	708	
Db	919	GGGGYHALAVCSGDSVWTTW-----GYNAPGQ	944	

RESULT 4

US-09-902-540-9932
; Sequence 9932, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

```

? APPLICANT: Wiegand, Roger C.
? TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
? FILE REFERENCE: 38-10(15849)B
? CURRENT APPLICATION NUMBER: US/09/902,540
? CURRENT FILING DATE: 2001-07-10
? PRIOR APPLICATION NUMBER: 60/217,883
? PRIOR FILING DATE: 2000-07-10
? NUMBER OF SEQ ID NOS: 16825
? SEQ ID NO 9932
? LENGTH: 1300
? TYPE: PRT
? ORGANISM: Myxococcus xanthus
? US-09-902-540-9932

```

Query Match	4.1%	Score 166.5	DB 4	Length 1300
Best Local Similarity	21.8%	Pred. No. 0.0018		
Matches 158	Conservative	95	Mismatches 240	Indels 231
Gaps				48
QY	24	FNEGAGLILYVRTDIGGMYRDAANGRIPLLD----	WVGW--NNMGYNGVSVIAADPIN	77
DB	639	YNPQPGMLPATIDLGTAYAGKTQVFRVGTDES	GGYTGMLLDLSEFNGI-----	689
QY	78	TNKVMAAVGMYTNSWDPNDAI LRSSDQATWQIT	PLPFLKGGNMPQGM--GGRL----	131
DB	690	TNTFFATIA-----PEDGYC-----	VNWPPIANAG--PDRSIAPGELSVLSYG	729
QY	132	AVDPNDNINILYFCAPSGKGLWRSTDGATWSQ	MNTPEPDV---CTYIANPTDITTTGYOSDI	187
DB	730	SAADPEGQVTF-----	TTAQ--TSGPAVTLAGANTLNPSTFAPQVTES	771
QY	188	QGVVMAAFDKSSSSILGQASKITFVGVA	DPNNPVFWSRDGGATWQAVFCAPTGFPHKGVF	247
DB	772	TALVFTL--TVSDGVKSTDSVTTVVALPNNP-	-----FTVNAGLDGIVEERAEX	818
QY	248	DPVNHVLIATNSNTGGPYDSSGDVWKFSVT	SGTWTRISVPVSTDDTANDYFGYSGLTIDR	307
DB	819	-----TLGSGASDADGNA-----	LTYLMTQVSGTPV--AVKDY-----	849
QY	308	QHPNTIMVATQISWNPDTIIFRST--DGGATWT	-RIWDWTSYPNKRSLRYLVDISAEPWLT	364
DB	850	TTPTATFIAPVT-LDBSLVFLRLTVSGIATV	NDTIVTVTNANRA-----PIVSDTSA	903
QY	365	FGVQNPVPVPSKLGWMD-EAMAI	DPFNSDRMLYG--TGATLYATN-----	407
DB	904	FAA-----GTVTVTASA	VDP-DGDALSYWEQTGSGTVAINGADTSAISFATPV	951
QY	408	-----DLTKWSDGGQIHIA	PMVK-CLEBTA-VNDLISPPSGAPL	444
DB	952	PGSYEFTVTATDGSASAKAVPV	TIIDGSLPANSAPTVMAGIDATANAGDVVTLGSGAS-	101
QY	445	ISALGD-----LGG-----	FTHADVTAVPSTIFTSPVFTGTSVDYAEINPS	486
DB	1011	-DAEGDTUTYHWEQIGGTDVTLTGAD-TL	TES--FTAPSTAGSYTLGLI-LTVSDGTSTT	1063
QY	487	-----IIVRAGSFDPSSQPNDRH	VAFSTDGGKNWPFQSGEPGGVTGGTVAASADGR--	539
DB	1066	SDAVERVIAA--DFGANPS--NTAPEAD	AGESAIV-AEGATVTLNGT-ATDADGDTLVI	1111
QY	540	VWAPDGPQPVVYAVFGNSWNA	SGCVPANAIQRSDRVNPKTFYAL-SNG--TFYRSTDG	596
DB	1119	VMTQ-IGGTPV--TUSDASSIL	TPFTAPASS-----DPIUFLMWSDGATVVDVVTSI	1163
QY	597	GVTFOFVAAGLPSSGAVGMFHAVP	KGKDLWLAASSGLYHSTNGGS-----SWSA	652
DB	1169	SVTEENVA--PVASARAVL-----	SGNOTSATLDGSSSDANGDVLTVTRWTOVSGPN	1211
QY	653	SAVN	656	
DB	1219	ATIS	1222	

RESULT 5
US-08-410-784A-5


```

; Sequence 5, Application US/08410784A
; Patent No. 5912413
; GENERAL INFORMATION:
; APPLICANT: MYERS, ALAN M.
; APPLICANT: JAMES, MARTHA G.
; TITLE OF INVENTION: ISOLATION OF SUI, A STARCH DEBRANCHING
; TITLE OF INVENTION: ENZYME, THE PRODUCT OF THE MAIZE GENE
; TITLE OF INVENTION: SUGARY 1
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin and Hayes LLP
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,784A
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Heine, Ph.D., Holliday C.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: ISU-002XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-2290
; TELEFAX: 617-451-0313
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 772 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-410-784A-5

Query Match 4.0%; Score 162.5; DB 2; Length 772;
Best Local Similarity 19.8%; Pred. No. 0.0018;
Matches 162; Conservative 90; Mismatches 272; Indels 293; Gaps 42;

QY 51 WPLLDVGNWNNWYN---GVVS---IAADPINTKNWAAVGMVMTNSWDPNDGALLRSSD 104
Db 108 WGP--NWYASNGWKGSGVSDVDANGDRPNFKLLDPYAEBSQDP----- 155
QY 105 QGATWQITPLPFLKGGNNPGRGMRGLAVDPNNDILYFGAPSGKGLWRSTDSGATWSQM 164
Db 156 -----LNPSNQ-----GNVFASAHYRTTDSGI----- 178
QY 165 TNFPDVGYYIANPTDTTG-----YQSDTQGVVWVAFDKSSSLG----- 203
Db 179 --YAPKGVVLVPSTQSTGKTPKRAQKODVIEVHVRG-----FTBQDTSIPAQYRGTYG 231
QY 204 --QASKTIFVGA-----DNNPV-----FWSRDGGATWQA 232
Db 232 AGLKASYLASLGVTAVEFLPVQETQNDANDVVPNSDANQNTWGYNTENYFSPDRRYATNK 291
QY 233 VPGAPTG-FIPKGVFDPVNVHLYIATNTGTPYDGSQGVWKFVSTGTVTRISPFVST 291
Db 292 AAGGTAFQVQAFHAGIKVYMDV-----VYNHTAEGGTWSSDPTTAT 338

; Sequence 6
; US-09-248-796A-14123
; Sequence 14123, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14123
; LENGTH: 1060
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796A-14123

Query Match 4.0%; Score 159.5; DB 4; Length 1060;
Best Local Similarity 20.9%; Pred. No. 0.0045;
Matches 147; Conservative 68; Mismatches 267; Indels 223; Gaps 33;

QY 7 TWSNVAIGGGGVDGIVFNEGAPGILYVTRDTGG-----MYRWDAANGRW--IPL-LDWV 58
Db 142 TKYKVPAGYRPPVDAYI-----SATDVNSYTLSYANBYTCAGGYWQRAFFTLRWT 191
QY 59 GNNV--WGYNGVVSAAADPINTKNWAAVGMVMTNSWDPNDGALLRSSDQATWQI-TPLP 115
Db 192 GYRNSDAGSNGIVIVATRTTVDSTTAVT---TLPPDPN-----RDKTKTIELKPIP 241
QY 116 FK-----LGGNMPGRGMRGLAVDPNNDILYFGAPSGKGLWRST-DSGATW 161
```


Db 242 TTTTTSYVGVTSYLTAKTAP---IGETATVIV---DIPYHTTTTTSKWTGTTTSTTH 295
Qy 162 SONTNFPD---VGTIANPT-DTGYQSDIQGVVWVAFDKSSSSIGQASKTIFVGVADPN 217
Db 296 TNFTSDIDTVIQVPLPNTVTTTBYSQ-----SFATTTITGPPGNTDVTVLIRPPP 348
Qy 218 NPV-----FWSRD---GGATWQAVCPAGTGFIPHKGVFDPVHVL-----YIATSN 261
Db 349 NHTVTTTTEWSSESYTTSTFTAPPGTDSVI---IKPPNPTVTTTTEWSSESYTTTIV 404
Qy 262 GGPYDG-----SSGDVWKFVSVTGWTTRISPVST-----TAN 295
Db 405 TAPPGTDTVLIRPPNHTVTTTTEWSQSYTT-TTTLIAPPGTDSVIIRPPNPTVTT 463
Qy 296 DVFYGSGLT-----IDQHPNTIMVATQISW-----PDT 325
Db 464 EYWSQSYATTTTITAPPGETDTVLIRPPNHTVTTTTE--YMSQSYATTTTITAPPGETDT 521
Qy 326 IIFRSTDCGATWTRIDWTSYXNRSIRYVLDISABPMLTFGV---QNPVPVPSPKL--- 378
Db 522 VLIRPPNHTVTT---TEYWSQSYTTTTTTLIAPPGTDSVIIRPPNPTVTTTTEYSQ 576
Qy 379 -----GWMDEAMADPFNSDRMLYGTGATLYATNDLTWKDGGQIHIA----- 421
Db 577 SYATTTTITAPPGTDTVLIRPPNHTVTTSTEYWSQSYATTTTITAPAGTDSVIIRPP 636
Qy 422 -PMVKGLE-----ETAVNDLISPPSGAPLISALDGLGGTHADTVAVPSTTFTSPVFTTG 475
Db 637 NPTVTTTTEYSQSYTTTNTVTTAPPGG-----TDSVIIRPPNPTV 676
Qy 476 TSDVYAEALNPSIIVRAGSFDPSQPNDRHVAFTDGGKWFQSGSPGGVT----- 525
Db 677 TTEYWSQSYATTI-----TVTATPGTDTVLIRPPNPTVTTTTEYWSQ 721
Qy 526 --TGGTVAASADGSRFVWAPGPGQPVVYAVGFGNSWAASQGVPA 568
Db 722 YATTTTITAPPGTDTVLIRPPNPTVTTTTEYWSQSYATTTTITVA 766

RESULT 7
US-09-346-237-8
; Sequence 8, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Biesgaard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629,200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 774
; TYPE: PRT
; ORGANISM: Favobacterium odoratum
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(774)
; OTHER INFORMATION: Isoamylase
US-09-346-237-8
Query Match 3.9%; Score 158.5; DB 3; Length 774;
Best Local Similarity 18.8%; Pred. No. 0.0035;
Matches 158; Conservative 103; Mismatches 265; Indels 313; Gaps 44;
Qy 76 INTNKVAA-----VGMVYNSWPDNGAILRSSDQATWITPL 114
Db 35 INPNKLGAAVDATKANVTFKWSKATRIELYLYSTGSAEKAKVMTNSGGISVT-I 93

Qy 115 PFKLGNPGRGMGRERLAVDPNNDNLYFGAPSGKGLMRSTDSGATWSQMTNPPDVGTYY 174
Db 94 P---FSTLSGQGLG-----GTLVYGYRANGPNW---PYNASWTGSSSL----- 130
Qy 175 ANPTDTTGYQSDIQG-----VWVAFDKSSSSIGQASKTIFVGVADPNP 219
Db 131 -----GFISDVDAAGNRFNPNKLLSDPYALESHDPTTATM-----TNGS 170
Qy 220 VFWSRDGGATWQAV---PCAPTGF-----PHKGVFDPV---NHLVYIATSN 262
Db 171 IYAS---GATYRNIDSGSAPKGIVLAGDTQATGKPTRALKDDVVYEAHVRGLTWDTS 227
Qy 263 -----GPYDSSGDVWKFVSVTGWTTRISPV-----PSTDTANDYFGYSGLT 305
Db 228 ITAAVRYGTGKAGLKAALAAALGVTAEFLPVQETQNDTNDPSTSGSNYGYMTLNY 287
Qy 306 ---DRQ-----HNTIMVATQISWVPTTIIIFRSTDCGATWTRIW 341
Db 288 PAPDRRYAYDKTPGGPTREFKEMVKAFHDNGIKVLVD-----VYVNTTGEAGNSPT- 339
Qy 342 DMTSPNRSIR-----YVLDISAB-PWLTGCVQNPVPPVS-----PKLGMDEAM 385
Db 340 DKTTYNTITSFRGLDNPYYSLTADFQNSWDNTGVGNYNTRNTIAQNLIVDSLAYWRDKL 399
Qy 386 AIDPFNSDRMLYGTGATLYATNDLTWKDGGQIHIAPMVKGLEBETA VNDLIS--PP----- 439
Db 400 GVDGVRFDLASVLGNSCQHCFCFNDKWDAG-----NALNRIVAEELPPRPAT 445
Qy 440 ---SGAPLIS---ALG---DLGGF-----THADVTAVPSTTFTSPVFTTGTSDVYA 481
Db 446 GSGVDLIAEPWALIGNSYQVGGFPGSWAEWNGAYRVDVVRQAQNKLGSAVITTC----- 499
Qy 482 ELNPSIIVR-AGSFDPSQPNDRHVAFTDGGKWFQG---SEPGGVTTGTTVAASADGS 537
Db 500 ---QMATRFAGSSD-----LYGDDGRKPMHSVNFITAHDGFTLKDLYSCNSKN 545
Qy 538 RFVW--APGDPGQPVVYAVGFGNSW-----AASQ-----GVP-----ANQA 571
Db 546 NQVMPYGPSDGGE-----DNNNSWDQGGIADQKARNGMALMLLSAGVPMIVGGDEA 599
Qy 572 IRS-----DRVNPKTFF-----ALSNGTFFVSTD-G 596
Db 600 LASMNNPNYLDSSANLWNSRTTQNNFQSFASKAMIAFRKAHPALRPANFYSSVDNN 659
Qy 597 GVTFPQVAAGLPSSGAVGVNF-----HAVPGK-EGDLWLAASSGLYHSTNGSSWSAIT 649
Db 660 GNVMEQLRWFKPDGCVADATYFNDANNHAIAMRIDGSEFGDTASAIYVAHN---AWSA-- 714
Qy 650 GYSSAVNVGFKSAPGSSYPVAVFVGTIGGTVGAYRSDDCGTTWVLINDQHOYGNWGO 708
Db 715 -----QVNFPLPWPAGKGSWYRVTDTCGWAEGASQVQAPGSE-ALVGGENTAIGLCOR 766

RESULT 8

US-09-489-039A-9813
; Sequence 9813, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9813
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9813

Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-136-574A-44

Query Match 3.8%; Score 152.5; DB 3; Length 1751;
Best Local Similarity 18.7%; Pred. No. 0.029;
Matches 194; Conservative 101; Mismatches 349; Indels 351; Gaps 49;

QY 6 YTNVNAVGCGFVGVGIVNEGAGILYVRD-----ICGMTRW-----D 45
DB 233 YTAATGFTYGGFIDDLGW---AAVWLXIATNDSSYLTAKBELMSEYANGTNTWTQCWDD 289
QY 46 AANGRIPLLDVGNWNGYGVUSIAADPINTNKVAAVGM-YTNWDPNDGAILRSSD 104
DB 290 VRYGTILMLAKITGKEL--YKGAVERNLDHWTDRITYTPKGMAYLTGM-----GSLRYATT 343
QY 105 QG-----ATWQ-----ITPLPKLGNMFGP-----GMGERLAVDPNNNDI 140
DB 344 AALACVYADWGCDSNKKTYLNFAKSQIDVALGST--GRSFVVGFGTNYQHPHHRN- 400
QY 141 LYFGAPSGKGLWRSDSGATWSQMTNFPD-----VGTVIANTPTDTGYQSDIQGVVW-- 192
DB 401 -----AHSSWANSMKIPEYHRHILYGVGGPGSDSYNDIDITDYQNE 444
QY 193 VAFDKSSSLGQASKTIFVGVADPNPNVFWSDRGATWQAVPCAPGTFPHKGVDPVNH 252
DB 445 VACDYNAGIVGALAK-----MYQLYGGEPID--DFKAIETPTND 481
QY 253 VLVYAT--SNTGPGYDGGSDGVKFSVTSCTWTRISPVSTDTAN-DYF-----CY 300
DB 482 EIVFESKFGNSQOP---NTEVISIYNYRTGW-----PPRTVTKLSFYIDITELIQAGY 534
QY 301 SGLTIDRQHPNTIMVATQISWMPDITIIFRSTDGATWTRIDWTSYPNRSLRYL-DISA 359
DB 535 S-----PDVVKVDT-----YYIEGKISGPPVWD-----KRNRIYVVLVDFSG 572
QY 360 BPWLTFGVOPNFPVPSFKLGW-----MDEAMADPPNSDRMLYGTGATLATNDLTWDSG 415

DB 573 TK-----YPGGEVHEHKKQAQPKISVPOGYPWMDPTN-DPSYKGLTGLEKKNKVIAYDNN 626
QY 416 QGIHIAPMVKGLGETAVNDLISP-----PSGAP----- 443
DB 627 N-----LVWGLEPGAATSTPAPTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 680
QY 444 -----LISALGDL-----GGTHADVTAV-----P 463
DB 681 VLYKNETSASTGSRPWFKIVNGSGSSVDLSRVKIRYWTYVTDGDKPQSAVCDWAQIGAS 740
QY 464 STIFTSPTFTTGTGTS-VDYAELNPSIIVRAGSPDPSSQPNDRHVAFSIDGKN----- 514
DB 741 NVTFNFVKLLSSGVSGADY-YLEVFGSSGAGQLQPGKDTGDIQVRFENKNDNSNYNQADWS 799
QY 515 WFQ-----GSEPGGVTTGGTVAASADGSRFVWAPGDP----- 546
DB 800 WLQSMNTNYGENAKVTLYVDGVLVWGOEPGATPAPTSTATPTPTPTPTPTPTPTPTPTPTPT 859
QY 547 -----GQPVVYAVGFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSN-GT-- 589
DB 860 VSATPTPAPTASPVG-GSYWTPSESYGALKVWYVGNLSSPTNVLNPK--IKIENVGTGTA 916
QY 590 -----FYRSTGGVT-----FQP-----VAAGLPSSGA-----VG 614
DB 917 VDLRSRVKRYWTYIDGEATQSVSVASSINPAYIDVRFVKLGANAGGADYYVEIGFKSGAG 976
QY 615 VMFHAVPGKEGDIMLAASSGLYHSTNGGSSWSA-----ITGVSSAVNVGFKSAPGS 666
DB 977 VLAAGQSTKEIRLSIQKSGSYNQSDYSVRSANSYIENEKVTGYIDDLVWGREPGRNA 1036
QY 667 SYPAVFVVGTVGTGTA---YRSDDCGTT-----WVLINDDDHQYGNWGOAIT 711
DB 1037 QIKWYVANGNLGSMNTNVLNPKIKIENVGTTAVDLRSRVKRYWTYIDGEATQSVSVTSIN 1096
QY 712 GDHANLRVYIGTNGRGI-VYGDIG 735
DB 1097 PAYIDVKFVKLGANAGGADYYVEIG 1121

RESULT 11

US-09-408-020-4
; Sequence 4, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOEP.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-09-408-020-4

Query Match 3.8%; Score 152; DB 4; Length 3472;
Best Local Similarity 21.0%; Pred. No. 0.082;
Matches 172; Conservative 100; Mismatches 244; Indels 304; Gaps 50;

QY 96 DGAILESSDQATWQITP-----LPFKLGGNMPGRGMGERLAVDPNNDNILYFGAPSGKGL 151
DB 986 DGVTVFLSD-GPTGRVYPSYSLNIPFDISSAAPG-----GFVIVPGVSDIAFSA-DGRNM 1038
QY 152 WRSTDSGATWSQMTNFP-DVG-----TYIANP--TDTGYQSDIQGVVWVAF 195
DB 1039 LVADETGGIHRILYRSPYEIGTDFIKSSLGFEFVETFSAPRVQDLAGIAFSDHGMILAA 1098


```
QY 196 DKSSS-----SLGQASKTIFVGVADPNPNPVFWS-----RDGGATWQAVPG- 235
Db 1099 GSGSVHRYSLSPYAVSGAKYETAMIG-GSPGLEFSDDLGRMFVDPDAGSETAAVYGL 1157
QY 236 -APTGFIPHKGVDFDV-----NHVLYIATSNVTGP----- 264
Db 1158 AAPYGI-----GEAPLPPLFLGVGABEATLSPDGRHILV-----PGRFGLSQYSLFSTNL 1208
QY 265 -----YDSSGD--VWKFVSISGTWTRISPVPSDTAN-----DYFGVSG----- 302
Db 1209 ELCAEPRIIDGSCDCEDGIYAFB-SPGRGEGVSLAASITAADPGIGELHGFAGPMPAPV 1267
QY 303 ---LTIDROH-----PNTIMVAT---QISWMPDT-----IIFR 329
Db 1268 MEQVTLDSRGTLRLVLDRTVDVTVPYKMWVEDSDGQTTLANSTLLNRENSNILLER 1327
QY 330 STDGA-----TWTRIW---DWTSYPNR-----SLRYLVDISAEPMLTFGVQ 368
Db 1328 LDDAAAGKISYTSVPFRTWSSPPLGTGATRPHTLFGDVRLLDIYDASGD----- 1379
QY 369 PNPPVPSPK-LGWDEAMADPNSDRMLVGTGA-----TLYATNDLTWDSGGQIHIA 421
Db 1380 -----VPSFSGIEFSDGCRM-----FVTGIGTGGINIFTLSAPFDITLPHKSGSTNIG 1428
QY 422 PMVGLBEETA VNDLISPPSGAPL-----ISALGD-----LGGPTH-----ADVT-A 461
Db 1429 -----GL-----SVSDLAFANGNSLTVLDVGVLRVVALGDDYVVTGTYQKFRITLDTQG 1481
QY 462 VPSTIFTSP-----VFTTGTSDYA-----ELNPSIIVRAGSPDSS-----QP 500
Db 1482 IPNSIYTSPLGSLQSFVAYDDRIDLVLGSPNDISSTTEIIPYSLPRP---DPPTGMDPTP 1538
QY 501 NDRHVAFTDGGKWFQSGEPGGVTTGG-----TVAASADGSRV-----WARGDPG-- 547
Db 1539 DGRRMFLSTENGIDQYLLSEFPVAVTTSVFLRTIPIDGGAEGIRFVDRNGRLVFPAGDGI 1598
QY 548 --QPVVAVAGNSW--AASGVAPANAQIRSDRVNPKTFYALSNGTFYRSTDGGVTFQPV 603
Db 1599 QRHELIITPYGASTSLLETVRGV-----TDCGPGENPA 1631
QY 604 AAGLPSSGAVGMFHVPGKEGLWLAASGLYHSTNGSGSSWSAITGVSSAVNVVGFKGSA 663
Db 1632 AGEIRLAGT---FNASDN-----VQSPSGIEFSGDGTGMF--VTGFGAGVNEFSLSA 1679
QY 664 P-GSSYPAVFVVGITGG---VTGAYRSDDCGTTWVLINDD 699
Db 1680 PFDTTLFVHVVELHDIGGQPAVDLAFAPED--GRITLLLLAAD 1717
```

RESULT 12

```
US-09-346-237-4
; Sequence 4, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Search Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346.237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER FILING DATE: 1998-07-02
; EARLIER FILING DATE: 1998-07-02
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Pseudomonas amylocleromosa
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
```

```
; OTHER INFORMATION: Isoamylase
US-09-346-237-4
Query Match 3.8%; Score 151.5; DB 3; Length 776;
Best Local Similarity 20.7%; Pred. No. 0.011; Mismatches 274; Indels 291; Gaps 47;
Matches 171; Conservative 89;
QY 27 GAPGILYVRTDIGMYRWDAAANGRIWIPLLDDWVGNWNGYNGVVSIAADPINTKNVAAVVG 86
Db 96 GITGAVY-----YGYRANGPNWPYASNMGKSGAGFVSVDVANGDRFNENKLLLDY 147
QY 87 MYTNSWPNDCGAILRSSDQATWQITPLPFKLGMPGRCGERLAVDNPNNDILYFGAP 146
Db 148 AQEVSQDB-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182
QY 147 SGKGLMRSTDSGATWSQMTNFPDV--GTYIANPTD--TTGYQSDIOGVVWVAFDKSS--SS 201
Db 183 KGVVLVPSQTGTGKPTKRAQKDDVIYEVHVRGFTSQDTSIPAQYQYTYTGAGLKASYLAS 242
QY 202 LQQAASKTIFVGVAD-----PN---NPVPW-----SRDGGATWQAVPGAPT-G-F 240
Db 243 LG-VTAVEFLPVQETQNDANDVVPNSDANQNYWYMTENYFSPDRRYAYNKAAGGPTAEF 301
QY 241 IPHKGVPDPVNHVLYIATSNVTGGPYDSSGDVWKFVSISGTWTRISPVPSST----- 291
Db 302 QAMVQAEPHNAGIKVYMDV-----VYNHTAEGGTWTSSTPTATTIYSWRGLDN 348
QY 292 -----DTANDYF-GYSGLTIDROHPNTI---MVATOISWVPTDITIFRSTDCGATWTRIW 341
Db 349 ATYYELTSGNQFYDNTGIGANFNFTNTVAQNLIVDSLAYWANTM---GVGDGR-----F 400
QY 342 DWTIS-YNRSLRYVLDISAEPMLTFGVQPNPVPSPKLGWMDMAID-----PFSNDRML 396
Db 401 DLASVLGNSCLNGATASA-----PNCENGYNFDDAASNVAINRLREFTVTPAA 451
QY 397 YGTGATLYATNDLTWDSGGQIH-IAPMVKGLSE-----TAVNDLISPPSGAPLI 445
Db 452 GSGGLDLFA---EPWALGGSYQLGFGFGQWSEWNLFRDSLROAQNEL-----GSMTI 502
QY 446 SALDGLGFTHADVTAVPSTIFTSPVFTTGTISVDYAEINPSIIVR-----AGS 493
Db 503 YVIQANDFSGS-----SNLFQSSGRSPWNSINFIDVHDGTMTKDVKVYSCNGANNQAWP 556
QY 494 FDPSSQPNDRHVAFSTDGK-----NWFGSBBPGVTTGGTV---AASADGSRF-VWAPGD 545
Db 557 YGPS-----DGTSTNYSDQCMGAG---TGAVDORRAARTGMAFEMLSAGT 601
QY 546 P-----GQPVVYAVGFGN-----SWAASQGVAPANAQIRSDR-----V 577
Db 602 PLMQGGDEYLRTLQCNNAYNLDSSANMLTYSWTTDQ---SNFYTFQRLIAPRKAHPAL 658
QY 578 NPKTFYALSNGTFFYRSTDGGVTFQPVAGLPSSGAVG-----VMFHAVPGKEGD 626
Db 659 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYAINGPS-- 700
QY 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVVGFKGSA 665
Db 701 --LGDSNIIYVAINGWSSSVTTLTPAPPSTGTYVRYVTDCTDNDGASTFV-----APG 751
QY 666 SSYPAVFVVGITGGVTVGAYRSDCGTTWVLINDDQHQYGNWGOAI 710
Db 752 SE-----TLIGG-----AGTT-----YQCCGQSL 770
```

RESULT 13

```
5457037-3
; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO;CARRERA, PAOLO;CAMERINI, BARBARA;
; GALLI, GIULIANO;LUCCHESE, GIUSEPPE;GRANDI, GUIDO;DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
```


Db 675 --LGDSNSIYVAYNGWSSSVTFLLPAPPSTQWYRVTDTCNDWDGASTFV-----APG 725

Qy 666 SSYPAVFVGTIGGVTGAYRSDDCGTTWLVINDQHOYGNWGOAI 710

Db 726 SE-----TLIGG-----AGTT-----YGCCGQSL 744

RESULT 15

5457037-5

; Patent No. 5457037

; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;

; GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO

; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOMYLASE

; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME

; NUMBER OF SEQUENCES: 7

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/1,797

; FILING DATE: 08-JAN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 749,621

; FILING DATE: 19-AUG-1991

; APPLICATION NUMBER: 224,114

; FILING DATE: 25-JUL-1988

; SEQ ID NO:5:

; LENGTH: 751

5457037-5

Query Match 3.7%; Score 148.5; DB 6; Length 751;

Best Local Similarity 20.7%; Pred. No. 0.018;

Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;

Qy 27 GAPGILYVRDITGGMYRWDANGRWIPLLDVWCHNNWYGVVSIADPINTNKVWAAVG 86

Db 71 GITGAVY-----YGYRAWGNWPYASNWGKSGAGFVSDVDANGDRFNPKLLDPY 122

Qy 87 MYTNSWDPNDGAILRSSDQATWQITPLPKLGNMPCRMGERLAVDPNNNDNILYFCAP 146

Db 123 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 157

Qy 147 SGKGLMRSTDSGATWSQMTNPPDV--GTYYIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201

Db 158 KGVVLVFTSTQSTGKTPRAQKDDVIYVHVRGFTQDTSIPAQYRGTYYGAGLKASVLAS 217

Qy 202 LQASKTIFVGVAD-----PN---NPVFW-----SRDGGATWQAVPGATG--P 240

Db 218 LG-VTAVEFLVQETONDANDWVPNSDANQNYGTYMTENYFSPDRRYAYNKAAGGPTAEF 276

Qy 241 IPHKGVPDPVNHVLYIATSNTPGPGYDGGSGDVWKFVTSGTWTRISVPVST----- 291

Db 277 QAMVQAQHNAGIKYIMDV-----VYNHTAEGGTWTSDDPTTATIIYWRGLDN 323

Qy 292 -----DTANDYF--GYSGLTIDRHPNTI---MVATQISWMPDTIIFRSTDGATWTRIW 341

Db 324 TTYVELTSGNQYFYDNTGIGANFNTYNTVAQNLIIVDSLAWNTM---GVDGFR-----F 375

Qy 342 DWTS--YPNRSURYVLDISAEPLWTFGVOPNPVPSPLKGMWDEAMAD-----PNSDRML 396

Db 376 DLASVLGNSCLNGAYTASA-----PNCPNGGYNFADAASNVAINRIIREFTVRPAA 426

Qy 397 YGTGATLVATNDLTKWDSGGQIH-IAPNVKGLEE-----TAVNDLISPPSGAPLI 445

Db 427 GGSGLDLFA-----BFWATGGNSYQLGGPPQGWSEWNGLFRDSLROAQNEL-----GSMTI 477

Qy 446 SALGDLGGFTHADVTAVPSTIFTSVPVFTGTSDVYAEINLPSIIVR-----AGS 493

Db 478 YVTQDANDFGS-----SNLFQSGSRSPWNSINFIDVHDGWTLDKDYVSCGNANNSQAWP 531

Qy 494 FDPSSQNDRHVAFSTDGK-----NWFQSGSPGGVTTGGTV---AASADGSRF--VWAFGD 545

Db 532 YGPS-----DGGTSTNYSWDQMSAG---TGAAVDQRRRAARTGMAFEMLSAGT 576

Qy 546 P-----GQPVVYVAGFN-----SWAASQGVFANAQIRSDR-----V 577

Search completed: October 5, 2005, 08:04:30

Job time : 33.4312 secs

Db 577 PLMQGGDEYLRTLQCNNNAYNLDSANMLTYSWTTDQ-----SNFYTPAQRLIAFRKAHPAL 633

Qy 578 NPKTFYALSNCTFYRSTDDGGVTFOVAAAGLPSSGAVG-----VMEHAVPGKEGD 626

Db 634 RPSSWYSGSLTWY-----QP-----SGAVADSNYWNNTSNYAIAYAINGPS-- 675

Qy 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVVFGKSPAG 665

Db 676 --LGDSNSIYVAYNGWSSSVTFLLPAPPSTQWYRVTDTCNDWDGASTFV-----APG 726

Qy 666 SSYPAVFVGTIGGVTGAYRSDDCGTTWLVINDQHOYGNWGOAI 710

Db 727 SE-----TLIGG-----AGTT-----YGCCGQSL 745

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:52:23 ; Search time 93.9808 Seconds
(without alignments)
3268.602 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTQPTWSNVAIGGGGFVD.....YIGTNGRIVGVDIGGAPSG 740

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues

Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4036	100.0	740	11	US-09-917-376-3
2	4036	100.0	740	11	US-09-917-376-6
3	4036	100.0	740	14	US-10-155-400-3
4	4036	100.0	740	14	US-10-155-400-6
5	4036	100.0	957	11	US-09-917-376-1
6	4036	100.0	957	14	US-10-155-400-1
7	2429.5	60.2	882	14	US-10-156-761-9395
8	1680	41.6	726	11	US-09-917-376-7
9	1680	41.6	726	14	US-10-155-400-7
10	1579	39.1	838	15	US-10-420-191-2
11	1573.5	39.0	818	14	US-10-026-994-2

12	1405	34.8	739	14	US-10-156-761-10111	Sequence 10111, A
13	1117	27.7	789	15	US-10-395-241-14	Sequence 12, Appl
14	1117	27.7	812	15	US-10-395-241-12	Sequence 14, Appl
15	1114	27.6	826	15	US-10-395-241-18	Sequence 18, Appl
16	841.5	20.8	555	10	US-09-927-827-47	Sequence 47, Appl
17	238	5.9	2468	14	US-10-246-330-4	Sequence 4, Appl
18	238	5.9	2468	15	US-10-282-122A-66335	Sequence 66335, A
19	199.5	4.9	2435	15	US-10-282-122A-47453	Sequence 47453, A
20	193	4.8	1465	15	US-10-369-493-13955	Sequence 13955, A
21	181	4.5	1308	15	US-10-282-122A-44566	Sequence 44566, A
22	172.5	4.3	783	15	US-10-369-493-14105	Sequence 14105, A
23	170	4.2	1074	15	US-10-282-122A-50616	Sequence 50616, A
24	168.5	4.2	1439	15	US-10-282-122A-44567	Sequence 44567, A
25	167.5	4.2	1119	14	US-10-245-802-12	Sequence 12, Appl
26	165.5	4.1	613	15	US-10-369-493-11181	Sequence 11181, A
27	163.5	4.1	1289	15	US-10-282-122A-44901	Sequence 44901, A
28	162.5	4.0	772	18	US-10-179-562-39	Sequence 39, Appl
29	158.5	3.9	774	9	US-09-833-435A-8	Sequence 8, Appl
30	158.5	3.9	774	14	US-10-375-720-8	Sequence 8, Appl
31	157	3.9	350	14	US-10-156-761-9225	Sequence 9225, Ap
32	157	3.9	995	16	US-10-437-963-109146	Sequence 109146,
33	156.5	3.9	2204	15	US-10-282-122A-64364	Sequence 64364, A
34	156	3.9	2117	15	US-10-120-801-63	Sequence 63, Appl
35	155	3.8	1096	17	US-10-741-849-7179	Sequence 7179, Ap
36	152	3.8	827	15	US-10-282-122A-47312	Sequence 47312, A
37	152	3.8	1721	15	US-10-027-806-4	Sequence 4, Appl
38	152	3.8	3472	13	US-10-027-806-4	Sequence 4, Appl
39	152	3.8	3472	13	US-10-034-623-4	Sequence 4, Appl
40	152	3.8	3472	14	US-10-027-801-4	Sequence 4, Appl
41	152	3.8	3472	14	US-10-029-120-4	Sequence 4, Appl
42	151.5	3.8	776	9	US-09-833-435A-4	Sequence 4, Appl
43	151.5	3.8	776	14	US-10-375-720-4	Sequence 4, Appl
44	151.5	3.8	2319	17	US-10-831-070-26	Sequence 26, Appl
45	151	3.7	901	17	US-10-939-262-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-917-376-3
; Sequence 3, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURES:
; OTHER INFORMATION: Catalytic domain GH74
US-09-917-376-3

Query Match	100.0%	Score	4036	DB	11	Length	740
Best Local Similarity	100.0%	Pred. No.	4.7e-284				
Matches	740	Conservative	0	Mismatches	0	Indels	0
						Gaps	0
Qy	1	ATTQPTWSNVAIGGGGFVDGIVNEGAPGLIYRTDIGMYRDAANGRIPLLDWGVW	60				
Db	1	ATTQPTWSNVAIGGGGFVDGIVNEGAPGLIYRTDIGMYRDAANGRIPLLDWGVW	60				
Qy	61	NNNGYGVSTAAADPINTNKVAAVGMVYTNWDNDGAILRSSDQGWATQITPLPFKLG	120				

Db 61 NNWNGVSVVSTAADPINTNKWAAVGMVNTSWDNDGAILRSSDQATWQITPLPKLGG 120
QY 121 NMPGRGMRERLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180
Db 121 NMPGRGMRERLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNPNPVFWSRGGATWQAVPGAPTGF 240
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNPNPVFWSRGGATWQAVPGAPTGF 240
QY 241 IPHKGVFPVNVHLVIATSNTPGPGYDSSGDVWKFESVTSCTWTRISPVSTDTANDYFGY 300
Db 241 IPHKGVFPVNVHLVIATSNTPGPGYDSSGDVWKFESVTSCTWTRISPVSTDTANDYFGY 300
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGATWTRIWDWTSYPNRSRLRYLVDISAE 360
Db 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGATWTRIWDWTSYPNRSRLRYLVDISAE 360
QY 361 PMLTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI 420
Db 361 PMLTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI 420
QY 421 APWKVGLBEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIFTSPTTGTSDY 480
Db 421 APWKVGLBEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIFTSPTTGTSDY 480
QY 481 AELNPSIIVRAGSPDPSSQPNDRHVAFSTDGKKNWFQSGEGVTTGGTVAASADGSRFV 540
Db 481 AELNPSIIVRAGSPDPSSQPNDRHVAFSTDGKKNWFQSGEGVTTGGTVAASADGSRFV 540
QY 541 WAPDGPQPVVYAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSNGTFYRSTDDGVTF 600
Db 541 WAPDGPQPVVYAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSNGTFYRSTDDGVTF 600
QY 601 QPVAAGLPSSGAVGVWFMHVPKEGDLMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
Db 601 QPVAAGLPSSGAVGVWFMHVPKEGDLMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
QY 661 KSAPGSSYPAPVFWVGTIGVTVGAYSDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720
Db 661 KSAPGSSYPAPVFWVGTIGVTVGAYSDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720
QY 721 YIGTNGRGIVYGDIGGAPSG 740
Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 2
US-09-917-376-6
; Sequence 6, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40197.4US01
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 6
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-917-376-6

Query Match 100.0%; Score 4036; DB 11; Length 740;
Best Local Similarity 100.0%; Pred. No. 4.7e-284;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTQYPTWSNVAIGGGFVGDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
Db 1 ATTQYPTWSNVAIGGGFVGDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
QY 61 NNWNGVSVVSTAADPINTNKWAAVGMVNTSWDNDGAILRSSDQATWQITPLPKLGG 120
Db 61 NNWNGVSVVSTAADPINTNKWAAVGMVNTSWDNDGAILRSSDQATWQITPLPKLGG 120
QY 121 NMPGRGMRERLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180
Db 121 NMPGRGMRERLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNPNPVFWSRGGATWQAVPGAPTGF 240
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTI FVGADPNPNPVFWSRGGATWQAVPGAPTGF 240
QY 241 IPHKGVFPVNVHLVIATSNTPGPGYDSSGDVWKFESVTSCTWTRISPVSTDTANDYFGY 300
Db 241 IPHKGVFPVNVHLVIATSNTPGPGYDSSGDVWKFESVTSCTWTRISPVSTDTANDYFGY 300
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGATWTRIWDWTSYPNRSRLRYLVDISAE 360
Db 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGATWTRIWDWTSYPNRSRLRYLVDISAE 360
QY 361 PMLTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI 420
Db 361 PMLTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTWKWDSGGQIHI 420
QY 421 APWKVGLBEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIFTSPTTGTSDY 480
Db 421 APWKVGLBEETAVNDLISPPSGAPLISALDGLGFFHADVTAVPSTIFTSPTTGTSDY 480
QY 481 AELNPSIIVRAGSPDPSSQPNDRHVAFSTDGKKNWFQSGEGVTTGGTVAASADGSRFV 540
Db 481 AELNPSIIVRAGSPDPSSQPNDRHVAFSTDGKKNWFQSGEGVTTGGTVAASADGSRFV 540
QY 541 WAPDGPQPVVYAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSNGTFYRSTDDGVTF 600
Db 541 WAPDGPQPVVYAVGFGNSWAASQGVAPANAQIRSDRVNPKTFYALNSNGTFYRSTDDGVTF 600
QY 601 QPVAAGLPSSGAVGVWFMHVPKEGDLMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
Db 601 QPVAAGLPSSGAVGVWFMHVPKEGDLMLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
QY 661 KSAPGSSYPAPVFWVGTIGVTVGAYSDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720
Db 661 KSAPGSSYPAPVFWVGTIGVTVGAYSDCGTTWVLINDDQHOYGNWGOAITGDHANLRRV 720
QY 721 YIGTNGRGIVYGDIGGAPSG 740
Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 3
US-10-155-400-3
; Sequence 3, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus

FEATURE:
OTHER INFORMATION: Catalytic domain GH74
US-10-155-400-3

Query Match 100.0%; Score 4036; DB 14; Length 740;
Best Local Similarity 100.0%; Pred. No. 4.7e-284;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVG 60
DB 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVG 60
QY 61 NNWGYNGVVSIAADPINTNKVAAVGMTNSWDNDGAILRSSDQATWQITPLPFKLG 120
DB 61 NNWGYNGVVSIAADPINTNKVAAVGMTNSWDNDGAILRSSDQATWQITPLPFKLG 120
QY 121 NMPGRGMRGLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGYTIANPTDT 180
DB 121 NMPGRGMRGLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGYTIANPTDT 180
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPFWRSRGGATWQAVPGAPTGF 240
DB 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPFWRSRGGATWQAVPGAPTGF 240
QY 241 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDMVKFSVTSGTWTRISPVSTDTANDYFGY 300
DB 241 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDMVKFSVTSGTWTRISPVSTDTANDYFGY 300
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIIDWTSYPNRSLRYVLDISAE 360
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIIDWTSYPNRSLRYVLDISAE 360
QY 361 PWLTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
DB 361 PWLTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
QY 421 APWKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY 480
DB 421 APWKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY 480
QY 481 AELNPSIIVRAGSFPSSQPNDRHVAFTDGGKNWFGSEPGVTTGGTVAASADGSRFV 540
DB 481 AELNPSIIVRAGSFPSSQPNDRHVAFTDGGKNWFGSEPGVTTGGTVAASADGSRFV 540
QY 541 WAPDGPQPVVYAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600
DB 541 WAPDGPQPVVYAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600
QY 601 QPVAAGLPSSGAVGVNMFHVPKEGDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660
DB 601 QPVAAGLPSSGAVGVNMFHVPKEGDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660
QY 661 KSAPGSSYPAVFVGTIGVTCAYRSDCGTTWVLINDDQHOYGNWGQAITGDHANLRV 720
DB 661 KSAPGSSYPAVFVGTIGVTCAYRSDCGTTWVLINDDQHOYGNWGQAITGDHANLRV 720
QY 721 YIGTNGRGIVYGDIGGAPSG 740
DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 4
US-10-155-400-6
Sequence 6, Application US/10155400
Publication No. US20030108988A1
GENERAL INFORMATION:
APPLICANT: DING, SHI-YOU
APPLICANT: ADNEY, WILLIAM S.
APPLICANT: VINZANT, TODD B.
APPLICANT: HIMMEL, MICHAEL E.
TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
TITLE OF INVENTION: CELLULOLYTICUS
FILE REFERENCE: NREL 01-36A

CURRENT APPLICATION NUMBER: US/10/155,400
CURRENT FILING DATE: 2002-10-22
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 740
TYPE: PRT
ORGANISM: Acidothermus cellulolyticus
US-10-155-400-6

Query Match 100.0%; Score 4036; DB 14; Length 740;
Best Local Similarity 100.0%; Pred. No. 4.7e-284;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVG 60
DB 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVG 60
QY 61 NNWGYNGVVSIAADPINTNKVAAVGMTNSWDNDGAILRSSDQATWQITPLPFKLG 120
DB 61 NNWGYNGVVSIAADPINTNKVAAVGMTNSWDNDGAILRSSDQATWQITPLPFKLG 120
QY 121 NMPGRGMRGLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGYTIANPTDT 180
DB 121 NMPGRGMRGLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGYTIANPTDT 180
QY 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPFWRSRGGATWQAVPGAPTGF 240
DB 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPFWRSRGGATWQAVPGAPTGF 240
QY 241 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDMVKFSVTSGTWTRISPVSTDTANDYFGY 300
DB 241 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDMVKFSVTSGTWTRISPVSTDTANDYFGY 300
QY 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIIDWTSYPNRSLRYVLDISAE 360
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIIDWTSYPNRSLRYVLDISAE 360
QY 361 PWLTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
DB 361 PWLTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
QY 421 APWKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY 480
DB 421 APWKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDY 480
QY 481 AELNPSIIVRAGSFPSSQPNDRHVAFTDGGKNWFGSEPGVTTGGTVAASADGSRFV 540
DB 481 AELNPSIIVRAGSFPSSQPNDRHVAFTDGGKNWFGSEPGVTTGGTVAASADGSRFV 540
QY 541 WAPDGPQPVVYAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600
DB 541 WAPDGPQPVVYAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600
QY 601 QPVAAGLPSSGAVGVNMFHVPKEGDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660
DB 601 QPVAAGLPSSGAVGVNMFHVPKEGDLMLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660
QY 661 KSAPGSSYPAVFVGTIGVTCAYRSDCGTTWVLINDDQHOYGNWGQAITGDHANLRV 720
DB 661 KSAPGSSYPAVFVGTIGVTCAYRSDCGTTWVLINDDQHOYGNWGQAITGDHANLRV 720
QY 721 YIGTNGRGIVYGDIGGAPSG 740
DB 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 5
US-09-917-376-1
Sequence 1, Application US/09917376
Publication No. US2004003834A1
GENERAL INFORMATION:
APPLICANT: DING, SHI-YOU

APPLICANT: ADNEY, WILLIAM S.
APPLICANT: VINZANT, TODD B.
APPLICANT: HIMMEL, MICHAEL E.
TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
TITLE OF INVENTION: CELLULOYTICUS
FILE REFERENCE: 40197.4US01
CURRENT APPLICATION NUMBER: US/09/917,376
CURRENT FILING DATE: 2001-07-28
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 957
TYPE: PRT
ORGANISM: Acidothermus cellulolyticus
FEATURE:
NAME/KEY: MOD RES
LOCATION: (957)
OTHER INFORMATION: Any amino acid
US-09-917-376-1

Query Match 100.0%; Score 4036; DB 11; Length 957;
Best Local Similarity 100.0%; Pred. No. 6.6e-284;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
Db 47 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106
Qy 61 NNWYNGVVSAAADPINTNKVAAVGMVNTSWDNDGAILRSSDQATWQITPLPFKLG 120
Db 107 NNWYNGVVSAAADPINTNKVAAVGMVNTSWDNDGAILRSSDQATWQITPLPFKLG 166
Qy 121 NMPGRGGERLAVDPNDNINILYFGAPSGKGLWRSTDGATWSQMTNPPDVGYIANPTDT 180
Db 167 NMPGRGGERLAVDPNDNINILYFGAPSGKGLWRSTDGATWSQMTNPPDVGYIANPTDT 226
Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFVMSRDGGATWQAVPGAPTGF 240
Db 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFVMSRDGGATWQAVPGAPTGF 286
Qy 241 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSDTANDYFGY 300
Db 287 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSDTANDYFGY 346
Qy 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIWDWTSYPNRSRYVLDISAE 360
Db 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIWDWTSYPNRSRYVLDISAE 406
Qy 361 PWLTFGVQPNPVPSPKLGWMDAEMADPFPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420
Db 467 APWKGLEETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSVPVTTGTSVDY 526
Qy 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDCGKWFQSGEPGGYTTGTVAAASADGSRFV 540
Db 527 AELNPSIIVRAGSDPSSQPNDRHVAFTDCGKWFQSGEPGGYTTGTVAAASADGSRFV 586
Qy 541 WAPGDPGPVYVAVGFGNSWAAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600
Db 587 WAPGDPGPVYVAVGFGNSWAAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 646
Qy 601 QPVAAGLPSSGAVGMFHAVPKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 660
Db 647 QPVAAGLPSSGAVGMFHAVPKEGDLWLAASSGLYHSTNGSSWSAITGVSSAVNVGFG 706
Qy 661 KSAPGSSYPVAVVGTIGVTVGAYRSDDCGTTWVLINDDQHQYGNWGOAITGDHANLRV 720
Db 707 KSAPGSSYPVAVVGTIGVTVGAYRSDDCGTTWVLINDDQHQYGNWGOAITGDHANLRV 766
Qy 721 YIGTNGRGIVYDGGAPSG 740

Db 767 YIGTNGRGIVYDGGAPSG 786

RESULT 6
US-10-155-400-1
Sequence 1, Application US/10155400
Publication No. US20030108988A1
GENERAL INFORMATION:
APPLICANT: DING, SHI-YOU
APPLICANT: ADNEY, WILLIAM S.
APPLICANT: VINZANT, TODD B.
APPLICANT: HIMMEL, MICHAEL E.
TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
TITLE OF INVENTION: CELLULOYTICUS
FILE REFERENCE: NREL 01-36A
CURRENT APPLICATION NUMBER: US/10/155,400
CURRENT FILING DATE: 2002-10-22
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 957
TYPE: PRT
ORGANISM: Acidothermus cellulolyticus
FEATURE:
NAME/KEY: MOD RES
LOCATION: (957)
OTHER INFORMATION: Any amino acid
US-10-155-400-1

Query Match 100.0%; Score 4036; DB 14; Length 957;
Best Local Similarity 100.0%; Pred. No. 6.6e-284;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
Db 47 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106
Qy 61 NNWYNGVVSAAADPINTNKVAAVGMVNTSWDNDGAILRSSDQATWQITPLPFKLG 120
Db 107 NNWYNGVVSAAADPINTNKVAAVGMVNTSWDNDGAILRSSDQATWQITPLPFKLG 166
Qy 121 NMPGRGGERLAVDPNDNINILYFGAPSGKGLWRSTDGATWSQMTNPPDVGYIANPTDT 180
Db 167 NMPGRGGERLAVDPNDNINILYFGAPSGKGLWRSTDGATWSQMTNPPDVGYIANPTDT 226
Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFVMSRDGGATWQAVPGAPTGF 240
Db 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFVMSRDGGATWQAVPGAPTGF 286
Qy 241 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSDTANDYFGY 300
Db 287 IPHKGVPDPVNHVLIATSNTPGPGYDSSGDVWKFVSVTSGTWTRISPVPSDTANDYFGY 346
Qy 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIWDWTSYPNRSRYVLDISAE 360
Db 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIWDWTSYPNRSRYVLDISAE 406
Qy 361 PWLTFGVQPNPVPSPKLGWMDAEMADPFPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420
Db 407 PWLTFGVQPNPVPSPKLGWMDAEMADPFPNSDRMLYGTGATLYATNDLTWDSGGQIHI 466
Qy 421 APWKGLEETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSVPVTTGTSVDY 480
Db 467 APWKGLEETAANDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSVPVTTGTSVDY 526
Qy 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDCGKWFQSGEPGGYTTGTVAAASADGSRFV 540
Db 527 AELNPSIIVRAGSDPSSQPNDRHVAFTDCGKWFQSGEPGGYTTGTVAAASADGSRFV 586
Qy 541 WAPGDPGPVYVAVGFGNSWAAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600
Db 587 WAPGDPGPVYVAVGFGNSWAAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 646

QY 601 QVAAAGLSSGAVGVMPHAFVPGKGGDLWLAASSGLYHSTNGSSWSALTGYSSAVNVGFG 660
DB 647 QVAAAGLSSGAVGVMPHAFVPGKGGDLWLAASSGLYHSTNGSSWSALTGYSSAVNVGFG 706
QY 661 KSAFGSSYPVAVVGTGIGVTCGAYRSDDCGTTWVLINDQHQYGNWQAITGDHANLRRV 720
DB 707 KSAFGSSYPVAVVGTGIGVTCGAYRSDDCGTTWVLINDQHQYGNWQAITGDHANLRRV 766
QY 721 YIGTNGRGIVYDGGAPSG 740
DB 767 YIGTNGRGIVYDGGAPSG 786
RESULT 7
US-10-156-761-9395
; Sequence 9395, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9395
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9395
Query Match 60.2%; Score 2429.5; DB 14; Length 882;
Best Local Similarity 59.0%; Pred. No. 2.4e-167;
Matches 438; Conservative 111; Mismatches 174; Indels 19; Gaps 8;
QY 6 YTSNVAIGGGFVGGIVNEGAPGILYVYRTDGGMYRWDAAANGRIPLLDWVGNWGY 65
DB 39 YSNVAIGGGFVGGIVNEGAPGILYVYRTDGGMYRWDAAANGRIPLLDWVGNWGY 98
QY 66 NGVVSIAADPINTNKVAAVGYTNSWDPNDGAILRSSDQATWQITPPLPKLGGNMPGR 125
DB 99 TGVSLSASVDPNPKYAAVGYTNSWDPNGAVLRSGDRGASWQKTDLPFLKGGNMPGR 158
QY 126 GNGERLAVDPNDNLIYPGAGSKGLWRSTDGATWSQMTNPPDVCTYIANPTDTTGYOS 185
DB 159 GNGERLAVDPNRSVLYLGPAGSKGLWRSTDGATWSQMTNPPDVCTYIANPTDTTGYOS 218
QY 186 DTQGVVWAFDKSSSLGQAKTIFVGVADPNPVPFWSRGGATWQAVCAPGTFPHKG 245
DB 219 DNQGIWVVFDESGTSGSSTRIVYGVADKONSVYRSTDTAGATWRLAGQPTGHLAHKG 278
QY 246 VFDPVNHVLYATNSGTGGYDGSQVWKFVSVTSGTWTRISPVSPSTDTANDYFGSGLTI 305
DB 279 VLDAANGCLYLAISKGGYDGGKGLWRYTTKTGTWNI SPVAEAT---YVFGSGLTV 335
QY 306 DRQHPNTIMVATQISWNPDTIIPRSTDGATWTRIMWDTSPNRSRYLVLDISAEWLPF 365
DB 336 DRQHPCTVMATAYSSWNPDTQLPRSTDGATWTRIMWDTSPNRSRYLVLDISAEWLPF 395
QY 366 GVQNPFPVSPKLGWDEAWADPFPNSDRMLYGTGATLYATNDLTKWDSCGQIHIAPMVK 425
DB 396 GANPAPPEQTPLKGMWTELEIDPFPDSARMYGTGATVYGTDLNLTWDSGSQPTIKPMAR 455

QY 426 GLEETA VNDLISPPS-GAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDVDAELN 484
DB 456 GLEETA VNDLISPPS-GAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDVDAELN 515
QY 485 PSIIIVRAGSFDSSOPNDRHVAFSTDDGKNWFOGSEPGGVTTGGTVAASADGSRFVWAPG 544
DB 516 PGTVVVRVGNLD--SGP--HVAFPSTDNGANWAFAGADPSVGGGTVAASADGSRFVWAPG 570
QY 545 DPGQPWYAVGFGNSWAASQGVPAQAQRSDRVNPKTFYALSNGTFFYRSTDDGVTQFQVA 604
DB 571 GTG--VQYTTGFGTSWSASAGLPAGAVESDRVDPKTFYFGKSGRFFVSSDGGATPTASA 628
QY 605 A-GLPSSGAVGVMPHAFVPGKGGDLWLAASS-----GLYHSTNGSSWSALTGYSSAVNVG 658
DB 629 ATGLPSGDS--VRFKALPGTKGDIWLAGGASGAYGLMWHSTDCGAAFTKLATVDQADTIG 686
QY 659 FOKSAPGSSYPVAVVGTGIGVTCGAYRSDDCGTTWVLINDQHQYGNWQAITGDHANLR 718
DB 687 FGKAAATGASYQTLTYTSKIGGVGRGIFRSTDKGASWTRVNDDAHQWGTGAAITGDPRVYG 746
QY 719 RVYGTNGRGIVYDGGAPSG 740
DB 747 RVYGTNGRGIVYDGGAPSG 768
RESULT 8
US-09-917-376-7
; Sequence 7, Application US/09917376
; Publication No. US2004003834A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Aspergillus aculeatus
US-09-917-376-7
Query Match 41.6%; Score 1680; DB 11; Length 726;
Best Local Similarity 46.2%; Pred. No. 4.7e-113;
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;
QY 1 ATTQPTWNSVAI-GGGGFGVGVFVNEGAPGILYVYRTDGGMYRWDAAANGRIPLLDWVGNWGY 59
DB 1 AASQAVTWKVVVGGGGFTPGIVFNPSPKAGVAYARTDGGAYRLN--SDDTWTPLMDWVG 59
QY 60 ---WNNWGVNVSIAADPINTNKVAAVGYTNSWDPNDGAILRSSDQATWQITPPLPK 116
DB 60 NDTWHDW---GIDALATDPVDTDRVYVAVGYTNEWDPNVSGILRSTDDGTWETKLPF 116
QY 117 KLGNNPGRGMGRRLAVDPNDNLIYPGAPSGKGLWRSTDGATWSQMTNPPDVCTYIAN 176
DB 117 KVGNPNPGRGMGRRLAVDPNKNLIFGARSGLKWKSTDYGATWSNVTSFTWGTYFQD 176
QY 177 PTDTTGYSQSDIQGVVWVAFDKSSSLGQAKTIFVGVADPNPVPFWSRGGATWQAVPQA 236
DB 177 SSST--YTSDPVGVAMVTFDSTSGSGSATPRIFVGVADAGKSVFKSEDAATWAWVSGE 234
QY 237 PT-GFTPHKGVDPVNHVLYATNSGTGGYDGSQVWKFVSVTSGTWTRISPVSPSTDTAN 295
DB 235 PQYGFPLPHKGVLSPEEKTLYISYANGAGPYDGTNGTGHKYNITSGVWTDISP---TSLAS 291
QY 296 DYFGYSLGTLTDRQHPNTIMVATQISWNPDTIIPRSTDGATWTRIMWDTSPNRSRYLV 355

Db 292 TTYGGLSVLDQVPGTLMVAALNCWPDDELIFRSTDSGATWSPIEWNGVPSINYYYSY 351
Qy 356 DISAEPMLTFCVQNP-PVPSPLKGMDEMAIDPFNSDRMLYGTGATLYATNDLTWDS 414
Db 352 DISAPMLTQDTSIDQFPV--RVGMVEALAIIDPFNSHMLYGTGLTVYGGHDLTNWDS 408
Qy 415 GGQIHAPMKVLETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTFTSPVFTT 474
Db 409 KHNVTKSLAVGIEEMAVLGLITPPGAPALLSAVGDDGFFYHSDLDAAAPNQAHTPTTGT 468
Qy 475 GTSVDYAEALNPSIIVRAGSPSPQPNDRHVAFTSDGKNWFGQSEPGVTTGGTVAASA 534
Db 469 TNGIDYAGNKPSNIVRSGASD--DYPT--LALSNFGSTWYADYAASTSTGTGAVALSA 523
Qy 535 DGSFVWAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRST 594
Db 524 DGDVLLMSSTSGALVSKSQG--TLTAVSSLPSCGAVIASDKSNTVFGGSAGAIYVSK 580
Qy 595 DGGVTFQVAAAGLPSGAGVGMFHAVPGKEGDLMLAASSGLYHSTNGGSSWSAI-TGVSS 653
Db 581 NTATSFTKTVS-LGSSSTTVNAI-RAHPSIAGDVMASTDKGLWHSTDYGSTFTQIGSGVTA 638
Qy 654 AVNVFGKSPGSSYPVAVVGTGTCVTCAYRSDCGTTWVLINDQHQYGNWGOA--IT 711
Db 639 GWSFGFGKASSTGSYVVIYGFPTIDGAAGLKFSEADAGTNQVVISDASHGFGS-GSANVYN 697
Qy 712 GDHANLRVYIGTNGRGIVYGDIGAPSG 740
Db 698 GDLQTYGRVFRGHERPGHLLRSQREPAG 726

RESULT 9
US-10-155-400-7
; Sequence 7, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: CELULOLYTICUS
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 7
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Aspergillus aculeatus
US-10-155-400-7

Query Match 41.6%; Score 1680; DB 14; Length 726;
Best Local Similarity 46.2%; Pred. No. 4.7e-113;
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;
Qy 1 ATTQPYTWSNVAI-GGGGFDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVG 59
Db 1 AASQAYTKNVVTVGGGGFTPGIVFNPSAKGVAVARTDIGGAYRLN-SDDTWTPLMDWVG 59
Qy 60 ---NNWNGYGVVSIADPINTNKVAAVGMVYNSWDNDGAILRSSDQATQITPLPF 116
Db 60 NDTWHDW---GIDALATDPDTRVYVAVGMVYTNWEDPNVGSILRSTDQGDWTWTKLPP 116
Qy 117 KLGCMFGRGMRGLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSOMTNPDPVGTYIAN 176
Db 117 KVGCMFGRGMRGLAVDPNNKNSILYFGARSGHGLWKSTDYGATWSNVTSTWTGTYQD 176
Qy 177 PTDTTGYQSDIQGVVWVAFDKSSSLGQASKITIFGVADPNPNFWSRDGGATWQAVPGA 236
Db 177 SSST--YTSDDPVIAGVWTFDTSGSSGSAIPRIFGVADAGKSVFKSEADAGATWAVVSGE 234

Qy 237 PT-GFIPHKGVDPVNVHLYIATSNVTGPGYDQSSGDVWKFESVTSCTWTTRISPVSTDTAN 295
Db 235 PQYGLPHKGVSPBEKTLIYSYANGAGPYDGTNGTVHKYNTISGVWTDISP---TSLAS 291
Qy 236 DFGVSGLTIDRQHPNTIMVATQISWMPDITIIFRSTDDGATWTRIWDWTSYPNRSURYVL 355
Db 292 TTYGGLSVLDQVPGTLMVAALNCWPDDELIFRSTDSGATWSPIEWNGVPSINYYYSY 351
Qy 356 DISAEPMLTFCVQNP-PVPSPLKGMDEMAIDPFNSDRMLYGTGATLYATNDLTWDS 414
Db 352 DISAPMLTQDTSIDQFPV--RVGMVEALAIIDPFNSHMLYGTGLTVYGGHDLTNWDS 408
Qy 415 GGQIHAPMKVLETA VNDLISPPSGAPLISALGDLGGFTHADVAVPSTFTSPVFTT 474
Db 409 KHNVTKSLAVGIEEMAVLGLITPPGAPALLSAVGDDGFFYHSDLDAAAPNQAHTPTTGT 468
Qy 475 GTSVDYAEALNPSIIVRAGSPSPQPNDRHVAFTSDGKNWFGQSEPGVTTGGTVAASA 534
Db 469 TNGIDYAGNKPSNIVRSGASD--DYPT--LALSNFGSTWYADYAASTSTGTGAVALSA 523
Qy 535 DGSFVWAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALSNGTFYRST 594
Db 524 DGDVLLMSSTSGALVSKSQG--TLTAVSSLPSCGAVIASDKSNTVFGGSAGAIYVSK 580
Qy 595 DGGVTFQVAAAGLPSGAGVGMFHAVPGKEGDLMLAASSGLYHSTNGGSSWSAI-TGVSS 653
Db 581 NTATSFTKTVS-LGSSSTTVNAI-RAHPSIAGDVMASTDKGLWHSTDYGSTFTQIGSGVTA 638
Qy 654 AVNVFGKSPGSSYPVAVVGTGTCVTCAYRSDCGTTWVLINDQHQYGNWGOA--IT 711
Db 639 GWSFGFGKASSTGSYVVIYGFPTIDGAAGLKFSEADAGTNQVVISDASHGFGS-GSANVYN 697
Qy 712 GDHANLRVYIGTNGRGIVYGDIGAPSG 740
Db 698 GDLQTYGRVFRGHERPGHLLRSQREPAG 726

RESULT 10
US-10-420-191-2
; Sequence 2, Application US/10420191
; Publication No. US20040067569A1
; GENERAL INFORMATION:
; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.
; APPLICANT: Rev. Michael W.
; APPLICANT: Zaretsky, Elizabeth J.
; APPLICANT: Haas, Jeffrey A.
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids
; FILE REFERENCE: 10210.200-US
; CURRENT APPLICATION NUMBER: US/10/420,191
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: US 60/373,987
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 2
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-420-191-2

Query Match 39.1%; Score 1579; DB 15; Length 838;
Best Local Similarity 43.0%; Pred. No. 1.2e-105;
Matches 325; Conservative 123; Mismatches 264; Indels 44; Gaps 20;
Qy 6 YTWNSVAI-GGGGFDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWV---GW 60
Db 21 PSKKNVKGCGGGFFVPGIIFHPKTKGVAVARTDIGGLYRLN-ADDSWTATVDGIADNAGW 79
Qy 61 NNWNGYGVVSIADPINTNKVAAVGMVYNSWDNDGAILRSSDQATQITPLPFKLG 120
Db 80 HNW---GIDAVALDPQDDQKVYAAVGMVYNSWDPSNGAIRSRDRGATWSTNLPPKVG 136


```
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10111

Query Match      34.8%; Score 1405; DB 14; Length 739;
Best Local Similarity 40.8%; Pred. No. 4.4e-93;
Matches 308; Conservative 104; Mismatches 257; Indels 86; Gaps 21;

Qy 6 YTNVNAIAGGCGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWGNNWGY 65
Db 42 YRWRNAVIGTGFVTVGLFHPSPVRLAYARDIGAYRWDDRGARWTPDLHLGWDDNWL 101
Qy 66 NGVVSIAADPTNTNKVAAAGMYTNSWDPNDGAILRSSDQATWQITPLPFLKLGNNMGR 125
Db 102 LGVEMAVDPHPDRLLYLAQRYVGDW---AAFFVSEDRSQSFIIYESPPFMGANDMG 160
Qy 126 GMGERLAVDPNNNDILYFGAPSGKGLMRSTDSGATWSQMTNFPDVGTVIANPTDTTGYQ 185
Db 161 GAGERLLVDPKSSSSLGQASKTIIFVGVDAPNNP-----VFWSRDGGATWQAVCAPGPF 240
Qy 186 DIQGVVWVAFPKSSSSLGQASKTIIFVGVDAPNNP-----VFWSRDGGATWQAVCAPGPF 240
Db 208 SGQGVWF-----LVAAKRTVYAGWGDGSGTANLYRTAD-GTTGAVGPRPSGT 257
Qy 241 ---IPHKGVFDPVNHVLYIATSGPYDGGSGDVWKFVSTGWTWTRISPV---PSTDT 293
Db 258 SAKVPLRAAYDTHRELYVTVDGAPGCGQSGVSKLRTATGATGTEVTPVPKPGTTSBG 317
Qy 294 ANDYFGYSGLTIDRQHPNTIMVATQISWMPDTIIFRSTDDGATWTRIDWTSYPNRSLRY 353
Db 318 SADTFAYGVAVDAARRPCTLVSTNNRWADGTVFRSTDDGRTWTSKX-----AA 368
Qy 354 VLDISAEPLTFGVQPNPVPSPKLGWDEAMADPPNSDRMLYGTGATLYATWDLTKWD 413
Db 369 VEDVSETPFLDWGDD-----KPKFGMWIQALAVDPYDSQHVYGTGATLYGTRDLKRW- 421
Qy 414 SGGQHIHAPVKGLEETAVNDLISPPSG-APLISALGDLGGFTHADVTAVPST-1FTSPV 471
Db 422 -----APRIRGLSESAVRQLISPPVGEAHLISGLGDIGVMYHERLTASPSRGMAINPV 474
Qy 472 FTTGTSDVYAEINPISIIVRAGSFDPSSQPNDRHVAFTDGGKNW--FGSESPGGVTTGGT 529
Db 475 FGSATGLAAARPAVYVRTGWGDHNG-----AYSHDGGRTWAPFEAQPIAKDAFGP 528
Qy 530 VAASADGSRFWA-----PGDPQGVVYA-----VGFSGNWAASQGVPAANAQIRSDRVNPKT 581
Db 529 IATSDGCGTLLWSFVHWDG-----TTAAHRSTDNAGSWSVSPKPGATFVADPADPTR 583
Qy 582 FYA--LSNGTFRSTDDGVTFQPVAAAGLPSSGAGVGMFHAVPGKEGDLWLAAS-SGLYHS 638
Db 584 FYAVDFDNGTLYASTDSGRSTFARAGGLP-SGDQSKLVAAPGRSGDGLWLSAKMNGLYRS 642
Qy 639 TNGSSWSAITGVSSAVNVGFKSAPGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLIND 698
Db 643 TDGGDTFARIDSCWASYTLGFKAADGADYPAIQVGVSTETITAYVRSDDAARTWVRIND 702
Qy 699 DOHQVNGWQAITGTHANLRRVYIGTNGRGIYGD 733
Db 703 DAHQWNGWEAVVGDPRIGHRVYLATNGRGIQYGE 737

RESULT 13
US-10-395-241-14
; Sequence 14, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
```

```
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: JP 2002-83433
; FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Geotrichum sp. M128
US-10-395-241-14

Query Match      27.7%; Score 1117; DB 15; Length 789;
Best Local Similarity 35.0%; Pred. No. 3.8e-72;
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;

Qy 6 YTNVNAIAGGCGFVDGIVFNEGAPGILYVRTDIGMYRWDAANGRWIPLLDWGNNWGY 65
Db 4 YEFKRVNAIAGGCGYITGIVAHPKTKDLYARTDIGAYRWDAAGTAKWIPLNDFIEAQMNI 63
Qy 66 NGVVSIAADPTNTNKVAAAGMYT--NSWDPNDGAILRSSDQATWQITPLPFLKLGNNMGP 124
Db 64 MGTESIALDPNNPDRLLYLAQRYVGDW---AAFFVSEDRSQSFIIYESPPFMGANDMG 119
Qy 125 RMGERLAVDPNNNDILYFGAPSGKGLMRSTDSGATWSQMTNFPDVGTVIANPTDTTGYQ 184
Db 120 RNNGERLAVNPNFNSNEVVMGTGT-BGIWKSSDRAKTWNTVTSIPDAF-----TNGIGYT 172
Qy 185 SDIQGVVWVAFPKSSSSLGQASKTIIFVGVDAPNNPFWSRDGGATWQAVCAPGTFI--- 241
Db 173 S-----VIEDP-----ERNGTIYASATAPG--MYVTHDGGVSEVPVAGQSSMLNRT 218
Qy 242 -----PHKGVFDPVNHVLYIATSGPYDGGSGDVWKFVSTGWTWTRI- 285
Db 219 TCAPPDKKPAIAPQPMKVALTP--NFLYVTVADYPGPWGVTGFEVWRQNRITSGANDDIT 276
Qy 286 -----SPVPSDTTANDYFGYSGLTIDRQHPNTIMVATQISWMPDTI----IPRSTDDGAT 336
Db 277 PRVGNSSPAPYNNQTFPAGGCGLSVDAATPNRLVIT-LDRDPGPAALDSIYLSLDAGAT 335
Qy 337 WTRIDWTS-----YPNRSLRYVLDISAEPLTFGVQV-----NPPVSP---KLG 379
Db 336 WKDTQLSPPSLENGWGHPTNAARY-KDGTVPWMLDFNNGPQWGGYGAPHGPTGLTKFG 394
Qy 380 WMDEAMADPPNSDRMLYGTGATLYATWDLTK---WDSGGQIHLIAP---MVKGLEETA 431
Db 395 WMSAVLIDPPNPEHMYGTGATLYATWDLTLRVEKDW-----APSWYLOIDGLEENA 446
Qy 432 VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVTTGTSTVDYAEINPISIVRA 491
Db 447 ILSLRSPKSGAALLSGIGDISGMKHDDLTK-PQKMFQAPQFNLDSIDAAGNFPNVVRA 505
Qy 492 GSFDESSQPNDRHVAFTDGGKNW--FGSEPGGVTT---GGTVAASADGSRFWAP--G 544
Db 506 GSSGHEYDSACARGAYATDGGDAWTFPTCPFGMAASHYQGSTIAVDASGSGTVMWSTKLD 565
Qy 545 DRGQPVVYAVFGPNSWAASQGVPA-----NAQIRSDRVNPKTFYALSNGTFYRSRTDGGV 598
Db 566 EQASGFWYSHDYGKTWS-----VPAGDLKAQTANVLSDKVQDGTFFIATDGGKFFVSTDDGK 621
Qy 599 TFQPVAAAGLPSSGAGVGMFHAVPGKEGDLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN 656
Db 622 SYAAGKAGLVT--GTSLMPAVNPVWAGDVWVPEGGLFHSDFGASFRVGTANATLVS 679
Qy 657 VGFSGS-----APGSSYPVAVFVVGTT--IGGVTGAYRSDDCGTTWVLINDHQHQVNGWQA 710
Db 680 VGAPKSKSGKKASAPSAVFIWGTDPKPGSDIGLYRSDDNNGSTWTRVNDQEHNTSG-PTMI 738
Qy 711 TGDHANLRVYIGTNGRGIYVGD 734
Db 739 EADPKYGRVYIGTNGRGIYADL 762

RESULT 14
```


Qy	492	GSFDPSSQPNDRHVAFSTDGGKNW--FQSEPPGGVTT---GGTVAASADGSRFFWAP--G	544
Db	507	GSSGHEYDSACARGAYATDGDANTIPPTCPGNNASHYQGSTIAVDASGSQIVWSTKLD	566
Qy	545	DPGQPVVYAVGFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALSNGTIFYRSTDDGV	598
Db	567	EQASGPWYSHDYGKTWS-----VPAGDLKAQTAANVLSKVQDGTIFYATDGGKFFVSTDDGK	622
Qy	599	TFQPVAAAGLPSSGAVGVMFHAVPGKEGDLWL-AAASSGLYHSTNGGSSWSAI-TGVSSAVN	656
Db	623	SYAAKGAGLVT--GTSLMPAVNPWAGDVWVPVEGGGLFHSDFGASFTRVGTANALVS	680
Qy	657	VGFGK-----APGSSYPVAVFVVGTT--IGGVGTAYRSDDCGTTWVLINDQHQYGNWGOAI	710
Db	681	VGAPKSKSDGKKASAPSAVFINGTDKPGSDIGLYRSDDNGSTWTRVNDQEHNYSG-PTMI	739
Qy	711	TGDHANLRRVYIGTNGRGIVYGDI	734
Db	740	EADPKVYGRVYLGTVNGRGIVYADL	763

Search completed: October 5, 2005, 08:08:31
Job time : 97.9808 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 5, 2005, 07:09:01 ; Search time 9741.62 Seconds
(without alignments)
3680.789 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTPQYTSNVAIGGGFVD.....YIGTNGRCIVYDGGAPSG 740

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO spoil/US0917376/runat_04102005_164328_26897/app query.fasta_1.2524
-DB=GenEmbl -QFMT=fastcap -SURFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNIT8=bits -START=1 -END=-1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -NORM SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0917376 @CNC 1 1 15287 @runat_04102005_164328_26897 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELOPEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.ste.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	4036	100.0	2869	6	AX700058 Sequence
2	2429.5	60.2	299800	1	AP005028 Streptomy
3	2371	58.7	296500	1	AL939128 Streptomy
4	2063	51.1	2823	6	AX565635 Sequence

SUMMARIES

5	2061.5	51.1	3262	1	AF078038S1
6	2009	49.8	12732	1	AE007608 Clostridi
7	1982	49.1	2950	1	AJ585344 Clostridi
8	1753.5	43.4	2579	8	AB1292929 Agarius
9	1722	42.7	2745	8	AB015511 Aspergill
10	1620	40.1	3359	8	AY040839 Aspergill
11	1579	39.1	2517	8	AY281371 Hypocrea
12	1560.5	38.7	2724	8	AK110506 Oryza sat
13	1405	34.8	299550	1	AP005031 Streptomy
14	1162.5	28.8	14520	1	AE0011809 Xanthomon
15	1160.5	28.8	13278	1	AE001712 Thermotog
16	1145	28.4	11910	1	AE012276 Xanthomon
17	1117	27.7	2367	6	AX924431 Sequence
18	1117	27.7	2646	6	AX924429 Sequence
19	1117	27.7	2716	6	AB089343 Geotrichu
20	1114	27.6	2481	6	AX924435 Sequence
21	979	24.3	2755	8	AB116528 Geotrichu
22	460.5	11.4	590	8	AB1534359 Agarius
23	368.5	9.1	24208	1	AY281357S2 unculature
24	325	8.1	202301	1	AE017286 Desulfovi
25	271.5	6.7	47537	1	AY328853 Vibrio pa
26	268.5	6.7	301950	1	AP006570 Gloeobact
27	262	6.5	58128	7	AY357582 Burkholde
28	261	6.5	49575	7	AY328852 Vibrio pa
29	238	5.9	20342	1	AE004613 Pseudomon
30	237	5.9	110000	1	AE017282 03 Continuation (14 o
31	226.5	5.6	110000	1	AE017180_13 Continuation (14 o
32	218.5	5.4	7938	1	AE011079 Methanosa
33	214	5.3	5600	1	AF182518 Pseudomon
34	214	5.3	303226	1	AE016774 Pseudomon
35	206	5.1	34866	1	BX640426 Bordetell
36	205.5	5.1	8833	1	AY608695 Burkholde
37	204.5	5.1	110000	1	RMS591985_10 Continuation (11 o
38	204	5.1	110000	1	AP006618_08 Continuation (9 of
39	199.5	4.9	8722	1	AF546905 Mycobacte
40	198.5	4.9	349497	1	BX640440 Bordetell
41	198	4.9	110000	1	CP000011 01 Continuation (2 of
42	196	4.9	110000	1	BX571966_25 Continuation (26 o
43	193.5	4.8	348257	1	BX640425 Bordetell
44	193	4.8	303400	1	AP006578 Gloeobact
45	192.5	4.8	207212	2	AL671876 Mus muscu

ALIGNMENTS

RESULT 1
AX700058
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION AX700058
VERSION AX700058.1 GI:29536021
KEYWORDS
SOURCE Acidotherrmus cellulolyticus
ORGANISM Acidotherrmus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Frankineae; Acidotherrmaceae; Acidotherrmus.
REFERENCE 1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
TITLE Thermal tolerant avicelase from Acidotherrmus cellulolyticus
JOURNAL Patent: WO 03012090-A 2 13-PBB-2003;
Midwest Research Institute (US)
FEATURES
source 1. 2869
/organism="Acidotherrmus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN

Alignment Scores:
Pred. No.: 6.19e-198 Length: 2869
Score: 4036.00 Matches: 740
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-09-917-376-3 (1-740) x AX700058 (1-2869)			
QY	1	AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp	20
DB	139	GGCAGCATCTAGCCGTACACCTGGAGCAACGTGGCGATCGGGGGCGGCTTTGTCGAC	198
QY	21	GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly	40
DB	199	GGGATCGCTTCAATGAAGGTGCACCGGAATTTGTACGTGCGGACGACATCGGGGG	258
QY	41	MetTyrArgTTPAspAlaIleAsnGlyArgTTPileProLeuLeuAspTTPValGlyTTP	60
DB	259	ATGTATCGATGGATGGATGCCGCCAACGGCGGTGGATCTCTTCTGTGGATTTGGTGGGATGG	318
QY	61	AsnAsnTTPGlyTyrAsnGlyValValSerIleAlaIleAspProIleAsnThrAsnLys	80
DB	319	AACAATTTGGGGGTACACACGGGTGCTCAGCATTTGCGCAGACCCCGATCAATACATAACAG	378
QY	81	ValTTPAlaIleValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyAlaIleLeu	100
DB	379	GTAATGGCGCGCTCGGAATGTACACCAACAGCTGGACCCCAACACGACGAGCGGATCTCTC	438
QY	101	ArgSerSerAspGlnGlyAlaThrTTPGlnIleThrProLeuLeuProPheLysLeuGlyGly	120
DB	439	CGCTCGTCTGTATCAGGGCGCAACGTGGCAATTAACGCCCTTGCCGTTCAAGCTTGGCGGC	498
QY	121	AsnMetProGlyATGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle	140
DB	499	AACATGCCGGCGGTGGATGGCGCAGCGGCTTGGGTGGATCCAAACAAATGACAAACATT	558
QY	141	LeuTyrPheGlyAlaProSerGlyLysGlyLeuTTPArgSerThrAspSerGlyAlaThr	160
DB	559	CTGTATTTCCGGCCCGCAGCGCAAGGGCTCTGAGAAAGCACAGATTCGCGCGCGACC	618
QY	161	TTPSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr	180
DB	619	TGGTCCCAAGATGACAACTTTCCGACGCTAGGACGATACATTTGCAAAATCCCACTGACACG	678
QY	181	ThrGlyTyrGlnSerAspIleGlnGlyValValTTPValAlaPheAspLysSerSerSer	200
DB	679	ACCGGCTATCAGAGCGATATTCAAGCGCTCGTCTGGGTCTGTTTTCGACAAATCTTCGTCA	738
QY	201	SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal	220
DB	739	TCGCTCGGCAAGCGAGTAAGACCATTTTGTGGCGGTGGCGGATCCCAATAATCCGGTTC	798
QY	221	PheTTPSerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyValAlaProThrGlyPhe	240
DB	799	TTCTGGAGCAGACGCGCGCGCGACGTGGCAGCGGTGGCGGCTGGCGCGACCGGCTTC	858
QY	241	IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn	260
DB	859	ATCCCGCACAGGGCGTCTTTGACCCGGTCAACACGCTCTATATTGGCCACACGACAAAT	918
QY	261	ThrGlyGlyProTyrAspGlySerSerGlyAspValTTPLysPheSerValThrSerGly	280
DB	919	ACGGGTGGTTCGTATGACGGGAGCTCCGCGCAGCGTCTGGAAATTTCTCGGTGACCTCCGGG	978
QY	281	ThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr	300
DB	979	ACATGGACGCGAATCAGCCCGGTACTCTCGACGACACGCGCAACGACACTACTTTGGTTAC	1038
QY	301	SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer	320
DB	1039	AGCGGCCTCATTATCGACCGCCAGCACCCGAAACAGATAATGGTGGCAACCCAGATATCG	1098
QY	321	TTPTTPProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArgIle	340
DB	1099	TGGTGGCGGACACCATTAATCTTTTCGAGCACCGACGGCGGTGGACGTGGACGCGGATC	1158
QY	341	TTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu	360
DB	1159	TGGGATTTGGACGAGTTATCCCAATCGAAGCTTCCGATATGTCTTGCATTTCCGCGGAG	1218
QY	361	ProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTP	380
DB	1219	CCTTGGCTGACCTTCGGCGTACAGCCGAATCCTCCCGTACCAGTCCGAGTCCGGCTGG	1278
QY	381	MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly	400
DB	1279	ATGGATGAAGCGATGGCAATCGATCCGTTCAACTCTGATCGGATGCTCTACGGAACAGGC	1338
QY	401	AlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIle	420
DB	1339	GGCAGTTTGTACGCAACAAATGATCTCACGAAGTGGGACTCCGGCGGCCAGATTCAATC	1398
QY	421	AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer	440
DB	1399	GGCGCATGGTCAAAAGGATTGGAGGAGACGGCGGTAAACGATCTCATACGCCCGCGTCT	1458
QY	441	GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr	460
DB	1459	GGCGCCCGCTCATCAGCGCTCTCGGAGACCTCGCGGGCTTCAACCCACGCCGACGTTACT	1518
QY	461	AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr	480
DB	1519	GGCGTGCCTACGACGATCTTCACGTCACCGGTGTTTCACGACCGCACCGGTCGACTAT	1578
QY	481	AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro	500
DB	1579	GGGGAATTTGAATCCGCTCGATCATCTTCGCGCTGGAACTTCGATCATCGACGCCAACCG	1638
QY	501	AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTTPPheGlnGlySerGlu	520
DB	1639	AACGACAGCGACGTCGCGTCTTCGACACGCGCGGCAAGAACTGGTTCACAGCAGCGAA	1698
QY	521	ProGlyGlyValThrThrGlyGlyThrValAlaIleSerAlaAspGlySerArgPheVal	540
DB	1699	CCTGGCGGGGTGACACGCGGGCGCACCGTCCGCGCATCGGCCGACGGCTCTCGTTTCGTC	1758
QY	541	TTPAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTTP	560
DB	1759	TGGGCTCCCGCGCATCCCGTCCGCTGTGTGTATCGCAGTCCGATTTGCAACTCTCTGG	1818
QY	561	AlaAlaSerGlnGlyValProHisAlaGlnIleArgSerAspArgValAsnProLys	580
DB	1819	GCTGTCTCGCAAGGTGTTCCGCCAATGCCAGATCCGCTCAGACCGGGTGAATCCAAAG	1878
QY	581	ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe	600
DB	1879	ACTTTCTATGCCCTATCCAAATGGAACCTTCTATCGAGACACGACGCGCGGCTGACATTC	1938
QY	601	GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal	620
DB	1939	CAACCGCTCGCGCGCGTCTCCGAGACGCGTCCGCTCGGTGTCTATGTTCCACGCGGTG	1998
QY	621	ProGlyLysGlyAspLeuTTPLeuAlaIleSerSerSerGlyLeuTyrHisSerThrAsn	640
DB	1999	CCTGGAAGAAGAGCGATCTGTGGCTCGCTGCTATCGAGCGGGCTTTACCACTCAACCAAT	2058
QY	641	GlyGlySerSerTTPSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly	660
DB	2059	GGCGGACGAGTGGTCTGCAATACCGCGGTATCTCTCCGCGTGAACGTGGGATTTGGT	2118
QY	661	LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyVal	680
DB	2119	AGTCTGCGCGCGGTGCTCATACCCAGCGCTTTGTCTCGGCACGATCGGAGGGCTT	2178
QY	681	ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTTPValLeuIleAsnAspGln	700
DB	2179	ACGGGGGGGTACCGCTCCGACGACTGTGGACGACCTGGGTACTGATCAATGATGACCG	2238
QY	701	HisGlnTyrGlyAsnTTPGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal	720

Db 2239 CACCAATACGGAATTTGGGACAGCAATACCGGTGACCAACGCAATTTACGGCGGTG 2298
 721 TrrleGlyThrAsnGlyArgGlyIleValTyrglyAspIleGlyGlyAlaProSerGly 740
 2299 TACATAGCAGCAACGCGGTGGAATTGTATACGGGACATTTGGTGGTCCGCGTCCGGA 2358

RESULT 2
 AP005028
 LOCUS
 Streptomyces avermitilis genomic DNA, complete genome, section 8/30.
 ACCESSION AP005028
 VERSION AP005028.1
 KEYWORDS GI:29605366
 SOURCE
 ORGANISM
 Streptomyces avermitilis MA-4680
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1
 AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
 TITLE Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
 MEDLINE 21477403
 PUBMED 11572948

REFERENCE 2
 AUTHORS Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.
 TITLE Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
 JOURNAL Nat. Biotechnol. 21 (5), 526-531 (2003)
 MEDLINE 22608306
 PUBMED 12632562

REFERENCE 3
 AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
 COMMENT This work was done in collaboration with Haruo Ikeda (*1), Jun Ishikawa (*2), Akiharu Hanamoto (*3), Chigusa Takahashi (*3), Mayumi Shinose (*3), Hiroshi Horikawa (*4), Hidekazu Nakazawa (*4), Tomomi Osonoe (*4), Norihiro Kishida (*4), Hieashi Kikuchi (*4), Tadayoshi Shiba (*5), Yoshiyuki Sakaki (*6, *7), Masahira Hattori (*1, *7) and Satoshi Omura (*1, *3).
 Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.

REFERENCE 4
 AUTHORS *1 Kitasato Institute for Life Sciences, Kitasato University
 *2 National Institute of Infectious Diseases
 *3 The Kitasato Institute
 *4 National Institute of Technology and Evaluation
 *5 School of Science, Kitasato University
 *6 Institute of Medical Science, University of Tokyo
 *7 RIKEN, Genomic Sciences Center
 Following url is also available.
 http://avermilis.lis.kitasato-u.ac.jp.
 Location/Qualifiers
 1. .299800
 /organism="Streptomyces avermitilis MA-4680"
 /mol_type="genomic DNA"
 /strain="MA-4680"
 /db_xref="taxon:227882"
 /note="This strain is also named as strain: ATCC 31267, NCIM 12804 or NRRL 8165.--synonym: Streptomyces

FEATURES
 source
 1. .299800
 /organism="Streptomyces avermitilis MA-4680"
 /mol_type="genomic DNA"
 /strain="MA-4680"
 /db_xref="taxon:227882"
 /note="This strain is also named as strain: ATCC 31267, NCIM 12804 or NRRL 8165.--synonym: Streptomyces

CDS
 /note="SAV1722"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC69433.1"
 /db_xref="GI:29605367"
 /translation="MAGRALAVPCADLTLPMTSAESDGRVDRLLRPVSKMLGEE
 LPVAVTERGLAHRDRVALIDRTGKVASAKHRLMSRLITVTAQADNRAVMTSDG
 KELWSTDQVDEALSEIVGRPVFLTDAPHEALDRADPEDVLRDTRPPVSATGR1
 GGSGEGTFFDFPAPVHVMTSTILARIALAALPAGSAQAERYRPNVILRPLTRGTEND
 WIDRTLHLGTGLVLRVARTPCRAIPTLAHGTVPADTAALRVLAHNRVTPTESMGFQ
 PCAGVYAOVLRPRGIRLGDVFEWA"
 complement (1232..3406)
 /note="SAV1723"
 /codon_start=1
 /transl_table=11
 /product="putative mycodextranase"
 /protein_id="BAC69434.1"
 /db_xref="GI:29605368"
 /translation="MHSNLSRYVRMSAIGAAVALAAGLLVALAPAAHAAAGADLPPT
 SVEASATITGIGDHTQGIASASGRQAVRLGACQGVETVFRANAVSVNV
 PDQSGALNVYNGTKLAKTLAVTSKYSTDVTSWIAKATHHFFDNRLLGQVQPG
 DKIAVEASTVTVTVADFEQVAAATKAGAVSVTDKADPHFGQSDSTQAFRNAIAA
 AQGVVWTPGIDYRLTSSNGVQNTLQAGSWHVSVHTSRFIDQSSSGGVHLKDFP
 VTGEVTERVSNPNDPNSLGSVSGMWLQHLGLWLTGNNDLVVNNRILDT
 TADGLNNGARGVRVNRNPLRNGDDALAMWSLYSPDNTSSPESNTISQPNLANGIA
 IYGTDLAVNNLISDTNAGSGLAISNQKFLDPFPPLAGITTVSGNTLVAGMNFN
 WQHPGALRVDSYDIAEANSITGTTDPSYSAFESFGGGHGYATKRVTVDTGATV
 RNTGTVVQAESQGAAGSNVQATGGAAGIYNCYPYAGSGTFTLTDGGGNAGSSTW
 GDCSWAPOQGNPDPTNRNLAGRPATATGSDVYTPGKAVDGASTWESNTNAP
 PQAWTVLSSQAVRLVLPATAWQARTOTLSVQSGTDSAYSTVVASQGYRFD
 ATGNTATVLPSTNRLYLKLVNTANTGTPAAGPSEVAYLS"
 complement (3464..5065)
 /gene="agl1"
 /complement (3464..5065)
 /gene="agl1"
 /note="SAV1724"
 /codon_start=1
 /transl_table=11
 /product="putative alpha-glucosidase"
 /protein_id="BAC69435.1"
 /db_xref="GI:29605369"
 /translation="MCQPIPARIQKWRVAVIYQVYVRSFADGCGTGDLAGVRAR
 LPLVAGLVDAIWFNYSYPMKGGYDVADYVAVDPAPGTLAEAEKLIABARELGIR
 TLVDIVPNVSDQHPFRAALAGPGSPERALPHFRGGAHAGELPNDWPSPAGST
 EBPVTLQPDGWLHLFTPEQDNLNAHVAVROHEHDVLFWFGEVGVLPDTERFARYL
 AKDPDLPIVEGRAPHPFVDRLDHIYRSWRAIADBYGAVFVGEVLPDTERFARYL
 RDELHTAFNPSFLSCPWADARLTATIDETLAHPVGAATVLCNHDVTRTVRYG
 RTDGTDFAAKAGTFTDLALGTRRARAAALSLALPGSYLYLQGEELGLPEARIPRE
 SIQDMFHFSGHVDGRCRVPVWADAPHAGFGGEPMLPQPARWSAIAADIQAED
 RLSMSLYREALRLRPAYADGPTLWLPADGVLAFAFARTEGLMCMVNLADTPAGLPAHS
 RLLSGSLDDQGLRPQDTAAWLRI"
 complement (5097..5972)
 /note="SAV1725"
 /codon_start=1
 /transl_table=11
 /product="putative integral membrane protein"
 /protein_id="BAC69436.1"
 /db_xref="GI:29605370"
 /translation="MSTRILVSPILSRPRGKAAYWTVFGVLLPALAFIPVYVMV
 TGAMKSPDAARTPPSLVPQKHLGTYDAWDLQMLPAHLWNTVQATGAWLLQLVFC
 TAAAYALSRLKPVFGKLVGLGILATLMVPAQALVPKYLTVADLPLIHTSLNDPLGI
 WLPVAVANFLYLLKRFDDQIPRDVLEAAEIDGAGRILTLMSIVPMSRPVGVYSIF
 ALVAVQDFLWPLWVFSDDTKQPIVALVQLSQNIQLTVLIAAMVIASIPVAMFLV
 QRHLAGISAGSTKG"
 complement (5969..6925)
 /note="SAV1726"
 /codon_start=1
 /transl_table=11
 /product="putative integral membrane protein"
 /protein_id="BAC69437.1"

Qy	6	TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn	25
Db	175758	TACAGCTGGAAGAACCGCGCTCGACGGGGCGGCTTCGTCCCGGCATCGTCTTCAAC	175817
Qy	26	GluclyAlaProGlyIleLeuTyrValArgThrAspileGlyGlyMetTyrAtgTTrpAsp	45
Db	175818	CGTCCGAGAGAACCTTCGCTACGCCCGACCGACATCGCGCGCGCTACCGTGGGCC	175877
Qy	46	AlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrpValGlyTTrpAsnAsnTrpGlyTyr	65
Db	175878	GAGTCTCGAAGACCTGAGCGCGCTCTCGACTCGTCTCGCTGGAGCGACTGGGGCAC	175937
Qy	66	AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIysValTrpAlaAlaVal	85
Db	175938	ACGGGTGTCTGTAGCTTCGCTCCGACTCCGTCACCGACCAACAGGTGTACCGCGCGCTC	175997
Qy	86	GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln	105
Db	175998	GGCAGGTACAGAACAGCTGGACCCGGGCAACGGTCCGTGTCAAGTCCGGCGACCGG	176057
Qy	106	GlyAlaThrTrpGlnIleThrProLeuProPheIysLeuGlyIysAsnMetProGlyArg	125
Db	176058	GGCCGAGCTGGCAGAGACCGACTCCCTTCAAGCTGGCGGGGAACATGCGCGGCGG	176117
Qy	126	GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla	145
Db	176118	GGCATGGCGAGCGCTCGCGTCCGCCAGAACAGGAACACGGTGTATCTCGCGCGC	176177
Qy	146	ProSerGlyIysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr	165
Db	176178	CCCAGCGCAAGGGGCTGTGCGCTCGACGGACTCGGGGGGCTCTCTGTCGACGCTCAC	176237
Qy	166	AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer	185
Db	176238	GACTTCCGAGACTCGGCACCTAGTCAGGACCGGACCCGACAGCGGGGTACGCTCC	176297
Qy	186	AspIleGlnGlyValValTrpValAlaPheAspIysSerSerSerLeuGlyGlnAla	205
Db	176298	GACAAACAGGGCATCGTGTGGTCACCTTCGACGAGTCGCGGGTCCGCGGGAGCTCC	176357
Qy	206	SerIysThrIlePheValGlyAlaAspProAsnAsnProValPheTrpSerArgAsp	225
Db	176358	ACGCGGACGGGTGACTGCGGGTCGCGCAAGAGCAACTCCGCTCTATCGCTCCACGGAC	176417
Qy	226	GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisIysGly	245
Db	176418	CGCGGCGGACTCGTTCGGCTGGCGCGCACCGGCCATCTCGCCACACAGGCG	176477
Qy	246	ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProThr	265
Db	176478	GTGCTGAGCGCGGACCGGCTGTCTACTCGCTACAGCAAGGCGGACCGGTAC	176537
Qy	266	AspGlySerSerGlyAspValTrpIysPheSerValThrSerGlyThrTrpThrArgIle	285
Db	176538	GACGCGCGAAGGACAGCTGTGCGGTATACAGCAGAGACCTGGAGCAACATC	176597
Qy	286	SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle	305
Db	176598	AGCCCGGTGCGGAGCGGCACAC-----TACTACGGCTTCAGCGGCTGACCGGTG	176648
Qy	306	AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr	325
Db	176649	GACCGGAGCATCCGGGACCGGTGATGGCGACTCGTACAGCTCTCTGTTGGCGGACAG	176708
Qy	326	IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer	345
Db	176709	CAGCTCTCCGCTCCACGACAGCGCGGACCTTGGACGAAGAGCTGGGACTACACCTCG	176768
Qy	346	TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPhe	365
Db	176769	TATCCGAGCGCTCGAACCGCTTCACCATGGATGTCGTCTCTCGCTCTGCTCACCTGG	176828

QY	366	GlyValGlnProAsnProValProSerProLysLeuGlyTyrMetAspGluAlaMet	385
Db	176829	GGAGCGAACCCGACCGCCGAGCAGACCCGAACTCGCTGGATGATCCGAGTCCCTG	176888
QY	386	AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla	405
Db	176889	GAGATCGACCGCTTCGACTCCGCGCGCATGATGATACGGAACGGCGCGCATCTACGCG	176948
QY	406	ThrAsnAspLeuThrLysTyrAspSerGlyGlyGlnIleHisIleAlaProMetValLys	425
Db	176949	ACGGACACCTTCGACCACTGGGACAGCGAAGCTTCCACATCAAGCCATGCGCGG	177008
QY	426	GlyLeuGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu	444
Db	177009	GGCTTGGAGGAGCGCGCTACACACTCGCTCGCTCCCTCCGCGCGGCCGACGCTG	177068
QY	445	IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer	464
Db	177069	TTACGCGCGCTCGGTGACATCGCGCGCTTCGCGCACCGGACCTCACACGCGTGCCTG	177128
QY	465	ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn	484
Db	177129	CTGATGTACGTCGCGCACTTCACACGAGCAGCAGCTCGACTACGCGCGGAGCCGAC	177188
QY	485	ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis	504
Db	177189	CCGGGACCGGTGCGGTGCGGCTCGCAATCTCGAC-----TCGGGTCCG-----CAT	177233
QY	505	ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyVal	524
Db	177234	GTGGGTTCTCGACGACGACACCGGCCCACTGTGTCGGGGGCGGACCTTCGCGGGT	177293
QY	525	ThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly	544
Db	177294	AGCGGGGTGGACGCTCGCGCCGCGTCCGACGCGAGTCGCTTCGTGTGAGCCCGCG	177353
QY	545	AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln	564
Db	177354	GGCACCGGG-----GTGCAGTACACGACCGCGGTTCGGCACCTCGTGTGCGCGTCCG	177407
QY	565	GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla	584
Db	177408	GGCTCCCGCGCGCGGCGATCGAGTCCGACCGCGGTCCGACCCGAGACCTTCTACGCG	177467
QY	585	LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla	604
Db	177468	TTCAAGTCGGGAGGTTCATCTAGTTCGAGCGCGGGCGGACCTTCACGCGTCCGCG	177527
QY	605	Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys	623
Db	177528	GCACGGGCTCGCGAGCGCGACAGC-----GTGCGCTTCAAGGCGCTGCGCGCGACG	177581
QY	624	GluGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHisSer	638
Db	177582	AAGGGGACATCTGCTGGCGCGCGCGGAGCGGCGCGGTACGCGCTGTGGCACTCG	177641
QY	639	ThrAsnGlyLysSerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly	658
Db	177642	ACGAGCGCGCGCGCGCTTCACCAAGCTCGCACCGTGCACGCGCGGCGCACCATCGCG	177701
QY	659	PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly	678
Db	177702	TTCCGCAAGCGCGCGCGCGCTCGTACCAAGCGCTCTACACGCGCGGAGATCGCG	177761
QY	679	GlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAsp	698
Db	177762	GGTGTGCGCGGATCTTCGCGGTCCGACCAAGGGGCGCGGTGGACCGCGTCAACGAC	177821
QY	699	AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg	718
Db	177822	GATGCCACAGTGGGGTTGGACGGCGCGCGGCGATCACCGGTGACCCGAGGTCTACGGG	177881
QY	719	ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro	738
Db	177882	CGCGGTATGTCTGCAGCAACCGCGCGCGGTATCTACGCGGCGGCTCTTCG	177941
QY	739	SerGly	740
Db	177942	GACGCG	177947
RESULT 3			
LOCUS	SCO939128	296500 bp	DNA linear BCT 11-FEB-2003
DEFINITION	Streptomyces coelicolor A3(2) complete genome; segment 25/29		
ACCESSION	AL031013	AL031013	AL031013
VERSION	AL031013.1	GI:24413907	
KEYWORDS	Streptomyces coelicolor A3(2)		
SOURCE	Streptomyces coelicolor A3(2)		
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
REFERENCE	1		
AUTHORS	Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, E., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrett, B.G., Parkhill, J., and Hopwood, D.A.		
TITLE	Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)		
JOURNAL	Nature	417 (6885), 141-147 (2002)	
MEDLINE	21964410		
PUBMED	12000953		
REFERENCE	2 (bases 1 to 296500)		
AUTHORS	Bentley, S.D.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk		
COMMENT	On or before Oct 26, 2002 this sequence version replaced gi:20520750, gi:20520810, gi:20520753, gi:20520666, gi:20520757, gi:20520758, gi:20520668, gi:20520759, gi:20520814, gi:20520765, gi:20520767.		
FEATURES	Location/Qualifiers		
source	1. .296500		
	/organism="Streptomyces coelicolor A3(2)"		
	/mol_type="genomic DNA"		
	/strain="A3(2)"		
	/db_xref="taxon:100226"		
gene	complement(144. .443)		
	/genes="SCO6480"		
	/notes="synonym: SC9C7.16c"		
CDS	complement(144. .443)		
	/gene="SCO6480"		
	/notes="SC9C7.16c, unknown, len: 99 aa"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein SC9C7.16c"		
	/protein_id="CAA22728.1"		
	/db_xref="GI:4154075"		
	/db_xref="UniProt/trEMBL:Q9ZBJ4"		
	/translation="MTYGCIAISMKTRPGFEEVSVLLLSADGLRDAGCELYVVLGSE DBETTWVTRVEDHDALEPAKDAIGRAMPLMTGFTKQEMSVAGGLG"		
	complement(517. .1563)		
gene	/gene="SCO6481"		
	/notes="synonym: SC9C7.17c"		
CDS	complement(517. .1563)		
	/gene="SCO6481"		
	/notes="SC9C7.17c, conserved hypothetical protein, len: 348aa; similar to several hypothetical proteins eg. SW108N MYCTU from Mycobacterium tuberculosis (340 aa) fasta scores; opt: 998, z-score: 1321.8, E(): 0, (47.68		

gene	identity in 355 aa overlap)."		gene	7111..8433	
	/codon_start=1			/gene="SC06485"	
	/transl_table=11			/note="synonym: SC9C7.21"	
	/product="conserved hypothetical protein"			7111..8433	
	/protein_id="CAA22729.1"			/gene="SC06485"	
	/db_xref="GI:4154076"			/note="SC9C7.21, unknown, len: 240 aa"	
	/db_xref="UniProt/TREMBL:Q92BJ3"			/codon_start=1	
	/translation="MRWENLTDSDHARAADTALFGADAVVTRTDTPTFRGITTPEHVR			/transl_table=11	
	ARSLNRVPGASMPFETVNPYRGCTHACVCFARKTHSLDIDTGLGPDQSIWVKV			/product="hypothetical protein SC9C7.21"	
	NAPVLRQLASRRWQGEHVMGTNVCYQACGRYRLMPCIIEALTERANPESILTK			CDS	/protein_id="CAA22733.1"
GTLIRLDLTRAARVTEVGLSVSGFTDAELMRTVEPFPAPERLEVRVLRGEG		/db_xref="GI:4154080"			
IGCVLMAPVPIFGDEPAQLRAITRAIAAGATSVTLVHLRPGAREWFMWIGRH		/db_xref="UniProt/TREMBL:Q92BJ1"			
HPHLYRRLYADGAYAPKWYQRRITRQVHDLAREYIGIGPARAGMPRVAEPHSRP		/db_xref="GOA:Q92BJ1"			
AEPAPPEVQLSLI"		/translation="MSVTSRYRDAMGEFWNEASGDRGAVFWDAEPTLTAGRHIAHPEP			
1800..2243		HLTDAALPLIDLCGNGTQTRYFAERFPHVVGADLSAALAHARHADPGGLAAYRQLD			
/gene="SC06482"		AADKAEALHAELGDANVYVRGLHQCEPDDQPLADLALLGERGRAFLVPEPSQA			
/note="synonym: SC9C7.18"		ACTVLMGLAQGPAGPPAKLAPVLRHGLTGPVDEAVPEYLRAGLTVLASGELPLVT			
1800..2243		gene	8603..9436		
/gene="SC06482"			/gene="SC06486"		
/note="synonym: SC9C7.19"			/note="synonym: SC9C7.22"		
1800..2243			8603..9436		
/gene="SC06483"			/gene="SC06486"		
/note="SC9C7.19, probable efflux protein, len: 815aa;			/note="SC9C7.22, possible transport associated protein,		
similar to many, several of which are involved in toxin			len: 277aa; similar to SW:DPFA_BACSU dbpa, dipeptide		
/antibiotic efflux eg. TR:O69070 (EMBL:AF061335)			transport protein (274 aa) fasta scores; opt: 448,		
/oxytetracycline exporter from Streptomyces rimosus (536			z-score: 604.3, E(): 2.3e-26, (32.3% identity in 269 aa		
aa) fasta scores; opt: 1020, z-score: 879.9, E(): 0,			CDS	overlap)."	
(36.9% identity in 542 aa overlap)."		/codon_start=1			
/codon_start=1		/transl_table=11			
/transl_table=11		Alignment Scores:			
/product="putative efflux protein"		Pred. No.:			
/protein_id="CAA22731.1"		2.37e-110			
/db_xref="GI:4154078"		Score:			
/db_xref="GOA:Q92BJ1"		2371.00			
/translation="MGGITLVQDPVRDPQVEFVVAQAAPGPPVPGDGRDAGLSP		Matches:			
RRVRLVFGMLLALLALEQMI VATAIPKIVGELHGLDRMSWAI TAYLLTSTVGLPI		431			
gene	YKGLDGLRKGQFOFALIVFVVGSAAGRAQTMQLIAFRAVQGVGAGLMIGVQAI		Percent Similarity:		
	IADI VPRERARYMGLIGASFGLASVAGPLGCGYFTDHLNWRWCYI NVPGFLVTMAV		72.29%		
	VANVLKLPRVPRKPLDILGLAASCTCLVLTSGWGTVEYMGSRVILGLAAGVA		Best Local Similarity:		
	ATLLFLFAPEPELIDRLFRDSIFNVTLVGLVGVGVALFGAASLYLPTFLQWMDGA		57.70%		
	SATESGLMLPMWGVGWASII SGLIISRTGTHYRHPILGSAISVVGWMLLSRLDADT		Query Match:		
	SRLHSI WMAVAGIGLVMPVL VLAQNSVTRPDLGTATANNYFRIGGSGVAHF		58.75%		
	GTLFAPVRLTDAALIPAEAGVGLPDAEAITPOLVHSLPPALRDAYIGIADAMPRIIF		IDB:		
	LYLVPVILGLIILFLAKEKPLVSHNAPVTPDPAHAPQPOPAQVPPARSPYAGGI		1		
	PVCVTQHPDGTVPRAALTLLDVGQOQIGRSGDDGRYALATPGSGAYVLI AAGG		US-09-917-376-3 (1-740) x SC0939128 (1-296500)		
	HQPAQVSPVPELVLDVILGGARLAGSVNTADGSPVRDAITVLTINVHGEVVAITR				
CDS	SGREGYVITELVAGEYTLAASAPAFRAALPVSOAARETRQDVLGAGVLRGTVR				
	ASGGRVADARVTLTLDAAAGNVVTLTTGPDGTFREVDLSGSSEYTVIAAGYPPVATVLQ				
	VAGGGRTERDLQLGHED"				
	4911..7493				
	/gene="SC06484"				
	/note="synonym: SC9C7.20"				
	4911..7493				
	/gene="SC06484"				
	/note="SC9C7.20, conserved hypothetical protein, len:				
	860aa; similar to many hypothetical proteins eqs.				


```
QY 6 TyrThrTrpSerAenValAlaIleGlyGlyGlyPheValAaspGlyIleValPheAen 25
Db TACACCTGGAGAACGCCCGCATCGACGGCGCGGTTCGTCGCCGCATCGTCTTCAAC 70380
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAasp 45
Db CGGACCGGAGAGGACCTGGCTAGCCCGGACCGACATCGAGGGCGCTACCGCTGGCAG 70320
QY 46 AlaAlaAenGlyArgTrpIleProLeuLeuAaspTrpValGlyTrpAenAasnTrpGlyTyr 65
Db GAGGAGTCCACACCTCGACCGCGCTCTCGACACCGCTCGCTGGAGCAGCTGGGGGCAC 70260
QY 66 AenglyValValSerIleAlaAalAaspProIleAenThrAenLysValTrpAlaAlaVal 85
Db ACAGCGGTGGTCTCGCTCTCGCTCGGACCGCTCGATCCGACCGCGGCTCTACGCGCGGTC 70200
QY 86 GlyMetTyrThrAenSerTrpAaspProAasnAaspGlyAlaIleLeuAtrGlySerAaspGln 105
Db GGCAGGTACACCAACGACTGGACCCGACCAACCGCGCGGTGCTCGCTTCGCGACCGG 70140
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAasnMetProGlyArg 125
Db GTGCGAGCTGGAGAACGGCGACTGCGGTTCAGCTGGCGGCAACATGCGCGCGCGC 70080
QY 126 GlyMetGlyGluArgLeuAlaValAaspProAasnAenAasnIleLeuTyrPheGlyAla 145
Db GGCATGGGTGAGCGCTGCGCGCTGACCGCACGACGACGACGCTGCTATACCTGGGCGCC 70020
QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAaspSerGlyAlaThrTrpSerGlnMetThr 165
Db CCCAGCGGCCACGGACTGTGGCGGTCCACCGCGCGCGCTACCTGTGCGAGTGAGC 69960
QY 166 AenProAaspValGlyThrTyrIleAlaAenProThrAaspThrGlyTyrGlnSer 185
Db GCTTCCCGAACCCCGGAATACCGCGAGGACCCGACGACGACGCTCGGCTACCGCTCC 69900
QY 186 AspIleGlnGlyValValTrpValAlaPheAaspLysSerSer---SerSerLeuGlyGln 204
Db GACACACGAGGCATCACCTGGGTACCTTCGACGAGTCGACGGCGCGCGCGCGCAGC 69840
QY 205 AlaSerLysThrIlePheValGlyValAlaAaspProAasnAenProValPheTrpSerArg 224
Db GGCACCGCGACCCCTACGTCGGGTTCGCGGACGAGAACGCGCTCTACCGCTCGACC 69780
QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 244
Db GACCGGGCGCGACCTGGGAGCGGCTCGCGGGCGAGCCGACGCGGTACCTGGCCCCACAG 69720
QY 245 GlyValPheAaspProValAasnHisValLeuTyrIleAlaThrSerAasnThrGlyGlyPro 264
Db GTGTGCTCGACCGGAGAACCGCTACCTGTACTCTCGCTACGACGACACCGCGCGCGC 69660
QY 265 TyrAaspGlySerSerGlyAaspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284
Db TACACGGCGGCAAGCGCGCTGTACCGGTACCGGACGCGCGACCGGACCTGGACGGAC 69600
QY 285 IleSerProValProSerThrAaspThrAlaAenAaspTyrPheGlyTyrSerGlyLeuThr 304
Db ATCAGCCCGCGCGGAGCGGACACCC-----TACTCGGCTTCAGCGCGCTCGAC 69549
QY 305 IleAaspArgGlnHisProAenThrIleMetValAlaThrGlnIleSerTrpTrpProAasp 324
Db GTGACCGGACGCGCGGGGACGGTGTGCGGACCGCTACAGCTCGTGGTGGCGGAC 69489
QY 325 ThrIlePheArgSerThrAaspGlyGlyAlaThrTrpThrArgIleTrpAaspTrpThr 344
Db ACGCAGATCTTCGCTCGACGACGACGCGCGGACATGGTCCGACGCGTGGAGCTACACC 69429
QY 345 SerTyrProAasnArgSerLeuArgTyrValLeuAaspIleSerAlaGluProTrpLeuThr 364
Db TCGTACCGGACCGCGAGAACCGCTACACGATGGACGCTCTCCTCGTCCCGCTCGCTCACC 69369
QY 365 PheGlyValGlnProAasnProValProSerProLysLeuGlyTyrTrpMetAaspGluAla 384
Db TGGGGCGGAACCCCGCGCGCCCGGAGCAGACCCCGAAGCTGGGTGGATGACGGAGGCT 69309
QY 385 MetAlaIleAaspProPheAasnSerAaspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404
Db CTGGAGATCGACCCGCTTCGACTCGGACCGGATGTATCGGCACCGGCGCCACGGTCTAC 69249
QY 405 AlaThrAenAaspLeuThrLysTrp---AaspSerGlyGlyGlnIleHisIleAlaProMet 423
Db GGCACGAGAACCTCAGCAACTGGGACGACGAGGGCGGACGCTTCGCCGCTCGAGCGGATG 69189
QY 424 VallysGlyLeuGluGluThrAlaValAenAaspLeuIleSerProProSerGlyAlaPro 443
Db GTGCGGGGCTTGAGAGACGCGCTCAACGACCTGGCCTCCCGCTCCCGCTCGGTGCCCG 69129
QY 444 LeuIleSerAlaLeuGlyAaspLeuGlyGlyPheThrHisAlaAaspValThrAlaValPro 463
Db CTGCTCAGCGCCCTGGGTGACGTGCGCGCTTCGCGCACACGAGCCTGACCGAGGTCCCG 69069
QY 464 SerThrIlePheThrSerProValPheThrThrGlyThrSerValAaspTyrAlaGluLeu 483
Db TCGATGATGTACACTCGCGCAACTTCACCTGACGACGAGCTGGACTTCGCGGAGACG 69009
QY 484 AenProSerIleIleValArgAlaGlySerPheAaspProSerSerGlnProAasnAaspArg 503
Db AMGCCGAGCTCGTGGTGGCGGGCAATCTCGAC-----TCCGGTCCG----- 68964
QY 504 HisValAlaPheSerThrAaspGlyLysAasnTrpPheGlnGlySerGluProGlyGly 523
Db CACATCGCGTCTCCAGGACAAACGGCGCAACTGGTTCGGGGCACACGACCGCTCCCGG 68904
QY 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAaspGlySerArgPheValTrpAlaPro 543
Db GTGACGGCGCGCGACGCTCGCGGGCGCGGACGCGCGCTTCGTGTGGAGCCG 68844
QY 544 GlyAaspProGlyGlnProValValTyrAlaValGlyPheGlyAasnSerTrpAlaAalSer 563
Db -----GAGGGTGGCGCGCTGACGACCGGCTTCGTACGTCCTGGCAGGCGCTCG 68790
QY 564 GlnGlyValProAlaAasnAlaGlnIleArgSerAaspArgValAasnProLysThrPheTyr 583
Db ACCGGCTCCCGCGCGCGGATCGTCGAGTCGACCGGCTGAACCCCGCGACCTTCTAC 68730
QY 584 AlaLeuSerAasnGlyThrPheTyrArgSerThrAaspGlyGlyValThrPheGlnProVal 603
Db GCGTTCACGTCGGGAGGTCTTACGTCAGTACGACGCGCGCGCGCGCTTCACCGCTCG 68670
QY 604 AlaAla---GlyLeuProSerSerGlyAlaValIglyValMetPheHisAlaValProGly 622
Db GCGCGACCGGGTGGCGCGCGGTGAC-----GGCGTTCGCTTCAGGCGCTCGCGGCG 68616
QY 623 LysGluGlyAaspLeuTrpLeuAlaAalSerSer-----GlyLeuTyrHis 637
Db -----GAGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68556
QY 638 SerThrAasnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAasnVal 657
Db TCGACGACCGCGCGCGGACCTTCACGAGCTCGCGGTGTCGACGCGCGCGGACACCGCTC 68496
QY 658 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIle 677
Db GCGTTCGCGAAGCGCGCGCGGAGCTGTAACAGACCTCTTCACGCGCGCGGATC 68436
QY 678 GlyGlyValThrGlyAlaTyrArgSerAaspCysGlyThrThrTrpValLeuIleAasn 697
Db -----GAGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68376
QY 698 AspAaspGlnHisGlnTyrGlyAasnTrpGlyGlnAlaIleThrGlyAaspHisAlaAasnLeu 717
Db -----GAGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68316
QY 718 ArgArgValTyrIleGlyThrAenGlyArgGlyIleValTyrGlyAaspIle----- 734
```


Db 68315 GCGCGGTGTACGTGGCGACCAACGCGCGGTGTACGTGGCGACACCTCGACACG 68256
 QY 735 ---GlyGlyAlaProSerGly 740
 Db 68255 GCGCGGGGCGACCGATCCGGGT 68235

RESULT 4
 LOCUS AX556535 2823 bp DNA linear PAT 29-NOV-2002
 DEFINITION Sequence 1 from Patent WO20077242.
 ACCESSION AX556535
 VERSION AX556535.1 GI:26000969
 KEYWORDS
 SOURCE unidentified
 ORGANISM unidentified
 unclassified.
 1
 Duffner, F. and Sjoeholm, C.
 Family 74 xyloglucanases
 TITLE Patent: WO 02077242-A 1 03-OCT-2002;
 JOURNAL Novozymes A/S (DK)
 FEATURES
 Location/Qualifiers
 1..2823
 /organism="unidentified"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"
 /note="Jonesia sp., DSM 14140"
 1..2823
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD58415.1"
 /db_xref="GI:26000970"
 /translation="MRURTVVATALTALLATAGGIAATSTPMVPTAPPSPSSPSGY
 SWSNVEIVGGFVFCIVFNQKDPGLVYARTDIGGAYRLNDSTGRMIPLTDHIGWDDWS
 HSGILSLATDPVDRNVRVLAQYTMQDQGLVYARTDIGGAYRLNDSTGRMIPLTDHIGWDDWS
 PGRGMLRLAVDPNNKVLVYFCAESGNLWKSITDYKWTGKVTSPFNAGNVPVGSITV
 YTGQNGVWVTPDTSKAGQTTQTIYGVADKONNVYSTDGGATWQRPVGGOPTFG
 LAQGLDKHQQLIATSDTGGPYDGSKGVDRWLDLISGGQWTRISPIPSISSNAPG
 YSLAIDRNPDIIMVVSQVSWNPDVYVSTDRGKTWSPILWELNGSOPRTKQYNHDI
 SGAPWLPDGNATKEPEANPKLGMWQTSFEIDPHNSDRFFYGTGAGIYGGTNLTWMDKG
 KKVDTIYKQAEIEETAAQDLAAPGNIGLYSALADIGGFTHKDI SQVPNKYYKNPH
 DTVTSIDPAESKPATVVRAGKSIGETTWSVGVSTDAGETWKPATGATPSGVKPGSITV
 SANASSIVAPEGAPRSTNSGWSMTSVGLPNAOVAASDRVNANTLYGFVDGKFEV
 STNGASPTAFAFTGPTSGNVRFRVPGRCQHLMLAGVSGSTYGMWRSYDGGKNT
 KVSQGEQDADFVGRVYSSGYPVIFTSKIDGVGRIPRSDDEKTKWRIINDNHNQWY
 WTGASITGDPVGRVYITGNRGII VGDSTPTPPQDSSGSGNTGATPTPETPDS
 DHNTGSGTGTGETSGPTSGSGTAPDSGNTSPGNTSSCKVRYSTDMGSGFTGAATI
 TNTSPSPINGWTLAFTYPSGQTISSVNSATQTLGSRNVVLKNSGNWPTIPAGGSVEFG
 FNGAYTSKNTPTSFSLNGVTCALAD"
 1..69
 70..2820
 /product="unnamed"

sig_peptide
 mat_peptide

ORIGIN
 Alignment Scores:
 Pred. No.: 7,686-97 Length: 2823
 Score: 2063.00 Matches: 386
 Percent Similarity: 67.21% Conservative: 112
 Best Local Similarity: 52.09% Mismatches: 225
 Query Match: 51.11% Indels: 18
 DB: 6 Gaps: 9

US-09-917-376-3 (1-740) x AX556535 (1-2823)

QY 6 TyrThrTrpSerAsnValAlaTleGlyGlyGlyGlyPheValAspGlyIleValPheAsn 25
 Db 130 TATAGCTGTGCAAGTTGAGATTGTGGCGCGCGGTGTGTGCGCTGGAATGTCTTTAAC 189

QY 26 GlyGlyAlaProGlyIleLeuThrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45
 Db 190 CAGAAAGACCCAGGGCTTGGTGTATGCCCGCACAGATATTGGTGGAGCGTACCGGCTCAAT 249

QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
 Db 250 GACTCCACCGCGGTGTGGATCCCTCTTACGGACCAATATTGGTTGGATGACTGGAGCCAC 309
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
 Db 310 TCCGGAATTCTCTCTGGCCACTGACCTGTGTGACCAACCGCGGTGACCTTGCAGCA 369
 QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
 Db 370 GCGAGTACTCATGCACTGGATCCCAAAACGGTGTCTATTTTGGTTCCGGGACAAA 429
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
 Db 430 GCGAGACCTGGAAAGACGATGCTGCGTTCCGTGTGGAGGCAACATGCCGGTCCG 489
 QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuThrPheGlyAla 145
 Db 490 GGGATGGGAGAACGCTCGCTTACCCCAAAATAATAAAGTCTCTGTACTTTGGTGG 549
 QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
 Db 550 GAATCGGGTAATGGAATATGGAAGTCACTGACTACGGGAAACCTGGGCAAGTGACG 609
 QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly---TyrGln 184
 Db 610 AGCTTCCCAATGCTGGTAACCTACGTCGCC-----GATGCTTCTGGTGGTACACC 660
 QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204
 Db 661 GGGCAAAACCAAGGTGTGTGTGGGTAACTTTTGACCCCAACATCAGCAAGCTGGTCA 720
 QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224
 Db 721 ACAACAAACCAATCTACGTAGTGTAGCAGCAACAGCAACACGCTTATTCATCATACC 780
 QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 244
 Db 781 GATGCGGGGCAACATGCAACCGCTCCTCGCAACCCACCGATTCTCTGGCGCAAAAG 840
 QY 245 GlyValPheAspProValAsnHisValLeuThrIleAlaThrSerAsnThrGlyGlyPro 264
 Db 841 GGAGTGTGACCATATAAAGGTGAGCAACCGCTCCTCGCAACCCACCGATTCTCTGGCGCCG 900
 QY 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284
 Db 901 TATGACGGTCCGAAGGGGATGTGGCGCTTACGATATTTCCAGTCCGCAAGTGGACTCGA 960
 QY 285 IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 304
 Db 961 ATCAGCCCAATTCCTCTTACC---TCATCGAATTCAGCGTTCGGGTACTCCGGTTAGCC 1017
 QY 305 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324
 Db 1018 ATTGACCGCAAAACCCCTGACCAACCATCATGTGTGTTCCTCCAGTGTCTTGGTGGCTGAC 1077
 QY 325 ThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspTrpThr 344
 Db 1078 ATGTAGCTATATATCGCTCCACCGCGGGGAAACCGTGGTCTATCTGGGAACGTGAAC 1137
 QY 345 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr 364
 Db 1138 GGTTCACAGCCCGGTACAAAGCAGTACCAACGATTTACTAGGGGCTCCTCGCTTGAC 1197
 QY 365 PheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAla 384
 Db 1198 TTTGGCAATACAGCCCAAGAGCTGAAGCTAACCCCAAACTCGGGTGGATGACCCAGTCC 1257
 QY 385 MetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404
 Db 1258 TTCGAGATTGACCGCATAACTCCGACCGCTTCTTCTATGGCAACCGCGCGGCTATCTAT 1317
 QY 405 AlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetVal 424


```
Db 1318 GCGGAACGACCTCAACAACTGGGATAAGGCAAAAGTTGATATCACTGCTCAAGGCC 1377
      ::::::::::::::::::::
Qy 425 LysGlyLeuGluThrAlaValAsnAspLeuSerProSerGlyAlaProLeu 444
      ::::::::::::::::::::
Db 1378 CAAGGATCGAAGAAACCGACACCGACCTTGCCTCCCTCGGAAACATTTGGCCTA 1437
      ::::::::::::::::::::
Qy 445 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 464
      ::::::::::::::::::::
Db 1438 TACTCAGCGCTGGCAGACATCGGTGGCTTACCCACAAAGATATCTCTCAGGTGCTTAAC 1497
      ::::::::::::::::::::
Qy 465 Thr---IlePheThrSerProValPheThrThrGlyThrSerValAspThrAlaGluLeu 483
      ::::::::::::::::::::
Db 1498 AAGTACTACTACAAGAACCCACATCAGACCGTCACTTCATGACTTCGCCGAAAGC 1557
      ::::::::::::::::::::
Qy 484 AsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503
      ::::::::::::::::::::
Db 1558 AAACGGCAACTGCTGCTGGTGCAGGCAAA-----TCTATTCTTGGGAAACCACTCA 1611
      ::::::::::::::::::::
Qy 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523
      ::::::::::::::::::::
Db 1612 TGGGTGGGTTCCTCAACTGATGCTGGCGAAACGTGGAAACCTGGCGCCACCATCAGGG 1671
      ::::::::::::::::::::
Qy 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543
      ::::::::::::::::::::
Db 1672 GTTAAAGCCCGCGCTCCATCACTGTCAGTAAATGCTTCATCAATGTGTGGGCCCG 1731
      ::::::::::::::::::::
Qy 544 GlyAspProGlyGlnProValValTrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563
      ::::::::::::::::::::
Db 1732 -----GAAGGTGGCGCTCCTCGAGTCCACCAACAGTGGCTCCTGCTGGTCAACGGTG 1785
      ::::::::::::::::::::
Qy 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583
      ::::::::::::::::::::
Db 1786 AGCGGGCTTCCTCAACATGCCAGTCCGATCAGACCGAGTCAACGCTAACTACTCTTTAC 1845
      ::::::::::::::::::::
Qy 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603
      ::::::::::::::::::::
Db 1846 GCCTTCGTTGACGGAAGTTCTATCACTCCACAAATGGTGGGGCTTCTCTTACCGCGTCA 1905
      ::::::::::::::::::::
Qy 604 Ala---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622
      ::::::::::::::::::::
Db 1906 GCTTTCACGGGGTTTCCACACAGTGA-----AAGTTCGTTTCCGTGGCGTGGCGGC 1959
      ::::::::::::::::::::
Qy 623 LysGlyGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHis 637
      ::::::::::::::::::::
Db 1960 CGTCAAGGTCACTCTGCTGCTGCTGAGTATCGGGAAGCACCTATGGGATGTGGCGG 2019
      ::::::::::::::::::::
Qy 638 SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657
      ::::::::::::::::::::
Db 2020 TCCACTGATGGCGGAAGAACTGGACCAAGGTATCTGGAGTTTCAGGAAGCGCATGCCGTT 2079
      ::::::::::::::::::::
Qy 658 GlyPheGlyLysSerAlaProGlySerSerTrpProAlaValPheValGlyThrIle 677
      ::::::::::::::::::::
Db 2080 GGTTCGGTAAGACCAACAGTAGTAGCGGATACCCGATGATTTTACATCCGCCAAGATT 2139
      ::::::::::::::::::::
Qy 678 GlyGlyValThrGlyAlaValArgSerAspAspCysGlyThrThrTrpValLeuIleAsn 697
      ::::::::::::::::::::
Db 2140 GATGGATTCGAGGAATTTCCGCTCTGACGATGAAGGCAAGACCTGGAAACGCTCAAC 2199
      ::::::::::::::::::::
Qy 698 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717
      ::::::::::::::::::::
Db 2200 GACAACCAACACAGTGGTATTGGACTGGCGCATCAATCACCAGTACCGGATGCGAA 2259
      ::::::::::::::::::::
Qy 718 ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAla 737
      ::::::::::::::::::::
Db 2260 GGGCGCGCTACTATTGGTACCAACGCTCGCGGCAATTATCTGTTGGTATTCCTCCACTCCC 2319
      ::::::::::::::::::::
Qy 738 Pro 738
      ::::::::::::::::::::
Db 2320 CCG 2322
      ::::::::::::::::::::
RESULT 5
AF078038S1
```

```
LOCUS AF078038S1 3262 bp DNA linear BCT 11-FEB-2001
DEFINITION Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 5 gene, partial cds.
ACCESSION AF078038
VERSION AF078038.1 GI:12743874
KEYWORDS 1 of 5
SEGMENT
SOURCE Caldicellulosiruptor sp. Tok7B.1
ORGANISM Caldicellulosiruptor sp. Tok7B.1
          Caldicellulosiruptor sp. Tok7B.1
          Bacteria; Firmicutes; Clostridia; Clostridiales;
          Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE 1 (bases 1 to 3262)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
          Williams,D.P. and Bergquist,P.L.
TITLE Multidomain and multifunctional glycosyl hydrolases from the
JOURNAL extreme thermophile Caldicellulosiruptor isolate Tok7B.1
MEDLINE Curr. Microbiol. 40 (5), 333-340 (2000)
PUBMED 20171169
10706665
REFERENCE 2 (bases 1 to 3262)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
          Williams,D.P. and Bergquist,P.L.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Biological Sciences, Macquarie University,
          Sydney, NSW 2109, Australia
FEATURES
    Location/Qualifiers
        1..3262
            /organism="Caldicellulosiruptor sp. Tok7B.1"
            /mol_type="genomic DNA"
            /isolate="Tok7B.1"
            /db_xref="taxon:80339"
        274..>3262
            /note="ORF5; multidomain"
            /codon_start=1
            /transl_table=1
            /product="glycosyl hydrolase 5"
            /protein_id="AAK06388.1"
            /db_xref="GI:12743879"
            /translation="MKRLISLPSILLINFLIAEFLQIGAKADTGIAPVYVKNVKI
EGGGFITGIVFNPKENLVYVIDIGGAYRSTDDGNTTQLMDVSDFNLLGVES
TATDPVDPKXVILACQTYTNSWTMGAILRSTDEGDTPEITPLPXLGGNMPXRL
GSRLAIDPNNRILYLGREGKWSKEDYGSWKVTSFNFCTYIEDPNCNDYLN
HITGVVVFDPDTPSGRPGCKIIYVGADKTSIYTKDGGQTQWALPQGTGLLPQ
RAKLSGDMLYITYSNTQCPNGDYCEVWRNTKTGEWKNISPMAAQDTFVGGGLAV
DAQNPKVNVAAALSSWPPDIYIWRSDGGSETKWCINWNGYPNRTLHYNDI SAAPWL
NFGXTBPTPEVSPKLGVMVGTLEIDFNSDXMLYGTGAXLYGCDLTLNWDKQNIIT
KYKAIETESVQALISPPVGPLFSAALGDIAGFRHEDLEKAPNWTYVQPNMGTTDI
DFAELNPNMVRVGNVDKQWNPNTNRIGFSYDGKSWFOGNTPEQGTSEGTVAAD
GSAVWVAPKGAQCYSTDNGNKWECANVPSEAIYVSDRVPNPKFYAFKNGKFYISAD
KGTFTESPAAGLPISGNFTVPGISGDIWLVGNMGWHSITDGYSFVKISGVEDAAS
IGFKPAREGETYPAIYIAKINGVRGIFRSDDCDKTWIRINDDKHQFGCANADITGDP
RVYGRVFNATNGIGIKWEIAYNISPSATPTSPSTPTPTPTATITPTATPTPT
PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
SAGSIRPWFKIVNGSSSVLSRVKIRYWTVDGKPSQSAVCDAQI CASNVTFVK
LSGYSKADYLYLEVGSAGQLQPKDGTGDIQVRFNKDWSNINQADDSWLSQMTD
YGENAKVTLYVDGVLVWGQPGGA"
        misc_feature
            274..2580
                /note="Region: endoglucanase domain"
                /evidence=experimental
        misc_feature
            2791..>3262
                /note="type IIIb cellulose binding site"
        misc_feature
            2581..2790
                /note="PT-linker"
ORIGIN
Alignment Scores: 1.08e-96 Length: 3262
Pred. No.: 2061.50 Matches: 387
Score: 65.62% Conservative: 96
Percent Similarity: 52.58% Mismatches: 238
Best Local Similarity: 51.08% Indels: 15
Query Match: 1 Gaps: 9
DB:
```


US-09-917-376-3 (1-740) x AF078038S1 (1-3262)

QY 4 GlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIle 22
DB 376 CAGCCATATGTTGGAGAAATGTTAAATTTGAAGTGGAGGAGTTTATCACTGGAATA 435
QY 23 ValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyr 42
DB 436 GTGTTTAAACCCGAAAGAGAAAACTCTGTTTACGTTAGGACAGATATAGGAGGCTTAC 495
QY 43 ArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsn 62
DB 496 CGTAGTACAGACGGTGGCAATACATGACACAACTTATGGACTGGTGTAGTTTTCATGAT 555
QY 63 TrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal--- 81
DB 556 TGGAAATTTGCTGGGTGAGAGCAATCGMACAGATCTGTGTGATCTTAATCGWGTATTA 615
QY 82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyValAlaIleLeuArg 101
DB 616 TTGGCTGCCAGGGAAGCTTATACCAATAGCTGGACAGATATCAATGGTGTCTATTGCGT 675
QY 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 121
DB 676 TCTACTGACGAAGGAGATACCTTTTGAATCACTCCACTCCCTTTTAAATTTGGTGGAAAC 735
QY 122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 141
DB 736 ATGCCRGAAGGAATCTGGGAAGAGGTGATCTATTTGATCTCCAAACAATAATAGAAATATA 795
QY 142 TyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyValAlaThrTrp 161
DB 796 TACTTAGAACACAGAGAAGGATGGTGTATGGAAAAGTGAAGATTATGGAGTGTCAATGG 855
QY 162 SerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 181
DB 856 AAAAAGTAACAAGTTTCTTAACCCAGCCACATATATTGAAGATCCAAACCTGCCCCCAAC 915
QY 182 GlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 201
DB 916 GATTATTTGAACCATATACAGGTGTGCTGGGTGATTTGATCTTACAAAGTGGTAGG 975
QY 202 LeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProValPhe 221
DB 976 CCTGTGAGGNAKACAGATATTTATGTGGTGTGTCAGATATAAACACACAGCACTAT 1035
QY 222 TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle 241
DB 1036 TATACAAAAGACGGTGGTCAGACATGGCAAGCATTTACCAGCACAGCCAACTGGCTTACTT 1095
QY 242 ProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThr 261
DB 1096 COTCAGCGTGCAAAATTAAGTTCT---GATGGGATGTTATATATATAACTTATAGTAACACT 1152
QY 262 GlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThr 281
DB 1153 CAGGGGCCATATAATGTTGATTATGGAAGATGAGATACATACTAAAGACCCGAGAA 1212
QY 282 TrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSer 301
DB 1213 TGGAAAAATATAAGTCCCAATGGCTGCTCAAGTACT-----TATTTTGGATATGGT 1263
QY 302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 321
DB 1264 GGATTTGGCTGTAGACCCGAGAAATCCGAAAGTAGTGGATGGTAGCAGCTCTTAGCTCATGG 1323
QY 322 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 341
DB 1324 TGGCCAGATACCTATATTATTTGGAGAAGTATCATGTTGGCGGAACCTGGAAATGTTATGG 1383
QY 342 AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 361
DB 1384 GNAUTGAATGGCTATCTCTTAATAGAACATTTGATTAACATGGAATATTTTTCGACGCTCT 1443

QY 362 TrpLeuThrPheGly---ValGlnProAsnProProValProSerProLysLeuGlyTrp 380
DB 1444 TGGCTAACTTTTGGAAWTACTTATCTACACCACTGAAGTAAGCCCAAACTTGGCTGG 1503
QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
DB 1504 ATGGTTGGAACACACTTGAATAATAGACCACTTAATTCAGATARAATGCTTTATGGAACAGGA 1563
QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
DB 1564 GCTAAWTGTTATGGATGCGATTTGACAAATTTGGGCAAAAGGGCAAAATATACCACT 1623
QY 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 440
DB 1624 AAAGTCAGGCTATAGGAATTTGAAGAACTTCAGTCCAGCACTGATAAGCCCAACAGTT 1683
QY 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460
DB 1684 GGACCACATTTATTCAGTGCCTGGGAGATATAGCGGATTCAGACATGAAGACTTGGAG 1743
QY 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
DB 1744 AAAGCACCTAATTTGGACATATGTTCCAGCTTAATATGGGACTACTACTGATATTGATTT 1803
QY 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500
DB 1804 GCTGATTTAATCCAAATTTTATGTTCTGTAGGGAATGTTGATAAACAGTGGGAATCCA 1863
QY 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly---Ser 519
DB 1864 AATACTAACAGAAATTTGTTCTCATATGATGTTGGAATAATCATGTTTCAGGGAATACT 1923
QY 520 GluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPhe 539
DB 1924 GAACCTCAAGGAACCAAGCGAAGTGTACAGTTGCAGCTGCTGCGAGATGGTAGTGCTGT 1983
QY 540 ValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer 559
DB 1984 GTTTGGGCACCA-----AAAGGCGCAAGATATGCTATTTCTACAGATAACGGAACA 2037
QY 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579
DB 2038 TGGGTTGAATGGCAAAATGTAACCTTCAGAAAGCTATTGTTATATTTCTGACAGAGTAATCCA 2097
QY 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 599
DB 2098 AATAAGTTTATGCAATTTAAGAAATGGCAAAATTTTATATAAGTGTCTGATAAAGGAAACA 2157
QY 600 Phe---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
DB 2158 TTTATCGAAAGTCTGCTGCTGGTTTACAAATATCAGGA-----AATTTCAAA 2205
QY 619 AlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer 638
DB 2206 ACTGTTCCAGGTATAGAGGTGATATCTGGCTAGTTGGTAAATATGGGATGTCGCACTCA 2265
QY 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658
DB 2266 ACTGATGTTGGATATTCATTTGTAATAATTTTCAAGGTTTGAGGATGTCAGCGAGTAGGA 2325
QY 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678
DB 2326 TTTGGTAAACCTGCTGGAAGGAGAAACATATCTCGCAATTTATCTTATGCAAAATAAT 2385
QY 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp 698
DB 2386 GGAGTAAGAGGTATATTTAGATCTGATCTGTGATAAACCTGGATGAAGATTAATGAC 2445
QY 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718
DB 2446 GACAAACATCAGTTTGGTTGTGCGAATGCAGACATAACTGAGAGATCCACCGGTGTACGGT 2505


```
RESULT 7
AJ585344
LOCUS
DEFINITION
2950 bp DNA linear BCT 11-OCT-2003
Clostridium thermocellum xgha gene for beta-1,4-xyloglucan
hydrolase, strain F7.
ACCESSION
AJ585344
VERSION
AJ585344.1
GI:37651952
KEYWORDS
beta-1,4-xyloglucan hydrolase; xgha gene.
SOURCE
Clostridium thermocellum
ORGANISM
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE
1
Zverlov, V.V.
New cellulosome components of C. thermocellum
Unpublished
2 (bases 1 to 2950)
Zverlov, V.V.
Direct Submission
Submitted (09-OCT-2003) Zverlov V.V., Fachgruppe f. Mikrobielle
Mikrobiologie, TU Muenchen, Am Hochanger 4, Freising, 85350,
GERMANY
FEATURES
source
1..2950
/organism="Clostridium thermocellum"
/mol_type="genomic DNA"
/strain="F7"
/db_xref="taxon:1515"
346..2887
/gene="xgha"
346..350
/gene="xgha"
359..2887
/gene="xgha"
/notes="belongs to GH74 and contains a cellulosomal
dockerin module"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="beta-1,4-xyloglucan hydrolase"
/protein_id="CAE51306.1"
/db_xref="GI:37651953"
tr\translation="MVKFTSKIAAFAAFAAATAIFGPAISQAQVTSVPYKWDNVV
IGGGFPMGPIVNETEKDLIYARADIGGAYRWDPTETWIPLLDHFQMDSESYVGE
STATDPVNRVTIVAGMYNDNLPMGAILRSTDRGETWEKTLIPKMGMPGRSM
IGVAVVDFKSSSTPGNPTKIYGVADKNESYRSDGGTGWKVPKQKGLLPHHG
VLAASGMLYITGDTGCTGPDGNGKGVQWKEKTRTGEWIDITPIYSSSDNRFCFAGLA
VDRONPDIIMVTSMAWMPDEYIFRSTGGATWKNIMWGMYPERILHYEDISAAPW
LDWGTSEQLPEINPKLGMIGDIEIDPPNSDRMYVTGATLYGCDNLTDWRGKVKI
EVKATGIECAVLDLVSPPEGAPLVSAVGLVGFVHDDLVKVPKMMVPSVSSGTGID
YAEVLPNFMAVLAKADLYDKKISFSDYDGRNMFQPPNEAPNSVGGGSVAADAASV
IWTPEASPAVTTDNGNSKVCNLTNGMAVAVSDRVNGKFPYAFNGKFPYISTDGLLT
FTDTKAPOLPKSVNKIKAVPGKEGHVMAAREGLMSTRTDGGYTFEKLISNVDYTHVVG
FGKAAEQDYMAIYITKIDNLVGFPSDDAGKTWRIINDEHGYGAVDTATITGDPVY
YGRYIATNGRGIYVEBPASDEPVPPTQVDDVGLVGLDNGNRINSTDLTLMKRYILK
SIEDLPVEDDLWAADNGDKINSTDTITLYLKQLLQAIPELPKK"
ORIGIN
Alignment Scores:
Pred. No.: 1,14e-92 Length: 2950
Score: 1982.00 Matches: 387
Percent Similarity: 65.81% Conservative: 100
Best Local Similarity: 52.30% Mismatch: 229
Query Match: 49.11% Indels: 24
DB: 1 Gaps: 14
US-09-917-376-3 (1-740) x AJ585344 (1-2950)
QY 2 ThrThrGlnProTyrThrTrpSerAsnValalaile---GlyGlyGlyPheValasp 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
458 ACCAGCGTGCCTTACAAATGGGACACAGTGTGTAATCGCGGCGGCGGAGGATTTATGCCG 517
```

```
381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1297 ATGATGGGAGACTTAAAGAAATTCATTCATTAATCGGATAGAAATGTTCTACGGTACAGT 1356
QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1357 GCACACTCTTATGGAACTGATGATTTGACTAATTCGGATAAAGTAAGAAATGACATC 1416
Db 421 AlaProMetValGlyGlyGluGluThrAlaValAsnAspLeuSerProPheSer 440
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1417 TCTGTAAAGGCAATGGAATTCGAAGATGTCAGTAATGATGCTGTGTAGTACCAACAAA 1476
QY 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaIleValThr 460
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1477 GCGGCACAACTTTTAAGCCTTTGTTGGAGATGACTCGCGATTTATCATGATGACATACT 1536
QY 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1537 AAGTTCCAGTAAGTAATGATGACACACCTAATCTTAGTCTACACAGCATTTGATTAT 1596
Db 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerGlnPro 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1597 GCAGAAAGTTCCTAACTTTGTTGTTGTTGGAATGTTGATAGTAAATAATCAA 1656
QY 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln---GlySer 519
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1657 CAAGATAAAGATTTGTGGAATTTCTACGATGTTGTTGAAGACTGTTCTCAGCTGGAAGC 1716
QY 520 GluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPhe 539
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1717 AATATATACAGGAGTTTATAAAGCAGGACTGTAGCAGCTGGTCTGCTGCTAAACTATA 1776
QY 540 ValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1777 GTATGCTCTCCA---GAGGAAGTGTCTAATGCAGCATATAGCACATAATGTAATAAA 1833
QY 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1834 TGGACACCGTGTCTGCTCTTCCACAGCGGAGCAAGAGTTAGATCAGACCGTGTAAATCCA 1893
QY 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 599
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1894 AAGAAATCTCATGTTTCTTAATAGTGTAAATCTATATAGTACTGATGCGAGTGCACAA 1953
QY 600 Phe---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1954 TTTTACTCAAGTTCACAAACAGTCTTCTCTACAAAGGCCAAA---GGAATC---TTTAAG 2007
QY 619 AlaValProGlyGlyGlyAspLeuTrpLeuAla---AlaSerSerGlyLeuTyrHis 637
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2008 ACTGTTATAGTTCATGAAGTGTATATTTGGATGCTGGAGGAAAGACGAGTATGCGCAT 2067
QY 638 SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2068 TCAACTGATCTCTGTCGCAACATTCACAAAAGTATCTGTTGTTGATGCTCTCTGATACAGTA 2127
QY 658 GlyPheGlyGlySerAlaProGlySerSerTyrProAlaValPheValGlyThrIle 677
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2128 GGTCTTGGAAATTCACAAACAGATGATGATATCCAGCTATATATACATGATGATCAAT 2187
QY 678 GlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsn 697
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2188 GATGGAACACGCGGAATATTCCTGTCAGATGATGAAGTGCACACTTGGGTAAAGATAAAT 2247
QY 698 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2248 GATGATGCTCACCATAATGTTTCCCGCAGACTACTGCTACTACTGAGATCTTAAACAAATAT 2307
QY 718 ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAla 737
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2308 GGACGAGTGTCTTGTGGAACATAATGAAGAGGTAATTTGTTTATGGAGACATAGATGGTCT 2367
```


QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyValArgThrAspIleGly 40
Db 518 GGTATAGTTTAAATGAAACGGAAGCGATTGATTTATGACGCTGCCGATATCGAGGA 577
QY 41 MetTyrArgTrpAspAlaAAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60
Db 578 GCGTACCGGTGGGATCCTTCGACCGAGACATGGATTCGGTTGCTGGACCATTTCCAAATG 637
QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAAspProIleAsnThrAsnLys 80
Db 638 GATGAGTACAGTTATACGGAGTGAAAGATATGCAACGACCCCTGTGGATCCGACCGT 697
QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
Db 698 GTTTACATAGTTGCAGGTATGTATACCAACGATTGGCTTCCTAATATGGAGCAATTCCT 757
QY 101 ArgSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheIysLeuGly 120
Db 758 CGCTCAACGGACAGGGGAAACATCGGAAACCACTACTGCCCTTCAAGATGGCGGA 817
QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
Db 818 AACATGCCGGGAAGATCCATGGGAGACGCTCTGGCATCGACCCCGAAATGACACAGGATT 877
QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
Db 878 CTTTATCTTGGACACAGATCGGAAACGGACTTTTGAGAGATACCGACTACGGTGAACA 937
QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 938 TGGTCCAAAGTTGAAAGTTCCCAAAATCCCGAACTTACATTTATGACCCG-----AAT 991
QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200
Db 992 TTTGATTATACCAAGACATATTATGGAGTAGTCTGGGTGTTTGTGCAAGAGACGACGT 1051
QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
Db 1052 ACACCGGGCAACCTCACCAAGACTATATGTTGGTGTGGCTGTGATGATAAACAAGAGTATT 1111
QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240
Db 1112 TACCGCAGTACGACGCGGGGTGTCCACTGGAAAGCAGTCCCGGACAACTTAAGGACTA 1171
QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
Db 1172 CTTCTCACCGGGGTTTGTG---GCATCCCAACGGAATGTTGTATATATTAACTTATCGTGCAT 1228
QY 261 ThrGlyGlyProTyrAspGlySer---SerGlyAspValTrpLysPheSerValThrSer 279
Db 1229 ACCTGCGGTTCCTTATGACGGACACGGAAGAGTACAGTTTGGAGTTCAATACACGTACA 1288
QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
Db 1289 GGGGAATGGATAGATATCACCCGATACCTTATTAACGACGAGTCACT---CGTTCTGCG 1345
QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
Db 1346 TTTGCAGGACTTCGATGCGACAGGAGAAATCTCTGACATTTAATGTTAACTTCCATGAAC 1405
QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyValAlaThrTrpThrArg 339
Db 1406 GCGTGGTGGCGGATGAATATATTTTCCGAGTACTGACGGCGGAGCTACATGGAGAAAT 1465
QY 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359
Db 1466 ATCTGGGAATGGGAATGATCCTGAAACGTATACCTGATTAATGAATAGATATTTCCGCA 1525
QY 360 GluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGly 379
Db 1526 GCACCGTGGCTGGATTGGGGAACGTGAGAAACAGCTGCCGGAATCAATCCGAAACTGGGA 1585
QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399

Db 1586 TGGATGATAGTGACATAGAGATTGACCGGTTTAATTCGACCGCATGATGTATGTACC 1645
QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 419
Db 1646 GGTGCAACTACTATGTTGGTGTGACAACTCTTACTGACTGGACAGAGCGCAAGTAAAA 1705
QY 420 IleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProPro 439
Db 1706 ATCGAGGTAAAGCTACCGGAATAGAAGAAATGCGGTATTAGACCTGGTAAAGCCCGCG 1765
QY 440 SerGlyAlaProLeuIleSerAlaIleuGlyAspLeuGlyGlyPheThrHisAlaAspVal 459
Db 1766 GAGGGTGCACCGCTTGTAAAGTCAGTTGGCGACCTTGTGCGTTTCTTATGATGACCTG 1825
QY 460 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 479
Db 1826 AAGTTGGTCCGAAAAAATGACGTT---CCTCTTATCTCTCAGTACGGGAATGTAT 1882
QY 480 TyrAlaGluLeuAsnPro-----SerIleValArgAlaGlySerPheAspPro 496
Db 1883 TATGCGGAGCTGTTCGAACTTTATGGCAATGTTGCAAGGCTGATTTGTATGATGTA 1942
QY 497 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe 516
Db 1943 -----AAGAAGATTTCTTCTCTTATGACGAGAGAAATTTGGTTC 1984
QY 517 Gln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAsp 535
Db 1985 CAGCACCTAATGAAGACCAAACTCGGTAGGCGCGGTTCGGTTCCGTTGCGGACCGAT 2044
QY 536 GlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGly 555
Db 2045 GCAAATCATTTATTTGGACACG---GAAATGCAAGTCTCTGCAGTT---ACACGAC 2098
QY 556 PheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAsp 575
Db 2099 AACGGAAACTCATGGAAGTTTGTACAAATCTTGGTATGGGTGCGGTGGTGCATCCGAC 2158
QY 576 ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp 595
Db 2159 CGTGTGAACGGTAAATAATTTCTACGCATTTCTATAACGGCAAAATCTTATATAAGCACGGAC 2218
QY 596 GlyGlyValThrPheGlnProVal---AlaAlaGlyLeuProSerSerGlyAlaValGly 614
Db 2219 GGTGATTHAACCTTTACCGATACAAAGGACCGCAGCTTCCCAAGTCGGTTAAC----- 2272
QY 615 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAla---SerSer 633
Db 2273 ---AAGATAAAAGCCGTACCGGGCAAGGAAGACATGATGGCTTGTCTGCAAGAGAAGGC 2329
QY 634 GlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSer 653
Db 2330 GGAATGTGGAGTCCCACTGACGCTGGATATACGTTTGAGAAACTCTCCCAATGTGACACA 2389
QY 654 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 673
Db 2390 GCTCATGTGGTAGGCTTCGGAAGGCGACCGGACAGGATTCATGCGCATTTACATT 2449
QY 674 ValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrp 693
Db 2450 ACCGGTAAATTTGACAACTGTTTAGGATCTTCCGTTCCGATGATGTCGCGCAAGACATGG 2509
QY 694 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 713
Db 2510 GTGCGTATCAACGACGACGACGATATGCGCTGTTGATATCTGCAATAACAGGTTGAC 2569
QY 714 HisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733
Db 2570 CCGAGAGTATACGGACGTGTATATATTCACCAACGGAAGAGGTATTTGTTTACGGCGAA 2629
RESULT 8
ABI292929


```
Db 128 CCCGGTATCGTCTTCAATCCCTCGCCAAAGGCGTGGCGTACGCGCGCACGGATATCGGA 187
Qy 40 GlyMetTyrArgTTPAspAlaAlaAsnGlyArgTTPileProLeuLeuAspTTPValGly 59
Db 188 GCGCGGTACCGTCTGAAC---TCGGACGACACGTCGACCCCGTGTATGGACTGGGTGGGC 244
Qy 60 -----TTPAsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProfile 76
Db 245 AATGATACCTGGCATGACTGG-----GGCATCGATGCGCTTGGCCACGGATCCGGTTC 295
Qy 77 AsnThrAsnLysValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAsp 96
Db 296 GACACGGATCGTGTACGTGGCGGTGGGCATGTATACCAATGAATGGACCCCAATGTG 355
Qy 97 GlyAlaIleLeuArgSerAspGlnGlyAlaThrTTPGlnIleThrProLeuProPhe 116
Db 356 GGATCGATCTCGCGTCGACGACGACGAGGTGATACCTGGACTCGACGACGAAGCTGCGGTTC 415
Qy 117 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn 136
Db 416 AAGGTGTGGGAATATCCCGGTCCGCGAATGGGTGAGAGATGGCGGTCCGACCCGAAC 475
Qy 137 AsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAsp 156
Db 476 AAGATAGCATTTCTCTATTGGACGCGGAGTGGGCACGGCTTGTGGAAGAGCAGAT 535
Qy 157 SerGlyAlaThrTTPSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn 176
Db 536 TATGAGCGACCTGTGTGAAATGTCACCTCTTCACGTGGACTGGAACGTATTTCCAGGAT 595
Qy 177 ProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTTPValAlaPheAsp 196
Db 596 TCGAGCTCGACC-----TACACTTCGGACCCCTGTGGGTATTGCTTGGGTGACCTTTGAT 649
Qy 197 LysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro 216
Db 650 TCGACCTCTGTGTAGCTCGGGGTCTGGACGCGCTAGGATCTTCTGTGGAGTTCGGGATGCT 709
Qy 217 AsnAsnProValPheTTPSerArgAspGlyGlyAlaThrTTPGlnAlaValProGlyAla 236
Db 710 GGCAGTCTGCTTCAAGTCGGAGGATGCGGGTGTACTTGGCTTGGGTGTCTGGAGAG 769
Qy 237 ProThr---GlyPheIleProHisLysGlyValAlaPheAspProValAsnHisValLeuTyr 255
Db 770 CCTCAGTATGGCTTCTCGGCACAAAGGGGTCTTGTCCCTGAGGAGAAGACGCTGTAC 829
Qy 256 IleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTTPlysPhe 275
Db 830 ATCTCTACGCAATGGCGCTGCTTACGATGGGCACAAACGGAACAGTCCACAAGTAC 889
Qy 276 SerValThrSerGlyThrTTPThrArgIleSerProValProSerThrAspThrAlaAsn 295
Db 890 AACATCACACGGAGTGGACGGATATAGTCCC-----ACCTCCCTTGGGAGT 940
Qy 296 AspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetVal 315
Db 941 ACCTACTACGGCTACGGTGTCTGTCGTGACCTCCAAAGTCTCGACACCTTGTATGGTG 1000
Qy 316 AlaThrGlnIleSerTTPProAspThrIleIlePheArgSerThrAspGlyGlyAla 335
Db 1001 GCTGCACCTCAACTGCTGGTGGCGGATGAGCTCATCTTCGTAGCACTGACTCTGGAGCG 1060
Qy 336 ThrTTPThrArgIleTTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeu 355
Db 1061 ACCTGTGCCCTATTGGAGTGAACGGTTATTCGAGCATCAATTACTACTACTACGCTAT 1120
Qy 356 AspIleSerAlaGluProTTPLeuThrPheGlyValGlnProAsnPro----ProValPro 374
Db 1121 GATATCAGCAACGCCCTTTGGATCCAGGACACGACTTCGACCCGACCAATTCGCCGTT--- 1177
Qy 375 SerProLysLeuGlyTTPMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394
Db 1178 -----CGGCTCGCTGGATGGTTGAAGCCCTTTGCCATTGACCATTTGACTCCCAACAC 1231
```

```
Qy 395 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSer 414
Db 1232 TGGCTCTACGGCACGGCCCTGACAGTATATGGCGGACACGACCTGACCAATTGGGACTCG 1291
Qy 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 434
Db 1292 AAGCACAAATGTACGGTCAAGTCACTCGCGGTGCGTATTGAGGAGATGGCAGTCTCGGA 1351
Qy 435 LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454
Db 1352 CTGATCACGCCCGCGGACCGCGCTGCTCTCGCTGTGGCGACGACGGCGGTTC 1411
Qy 455 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThr 474
Db 1412 TACCACCTCGGATCTGACGACGCGCCAAACCAAGCGTACACACGCGCACTACGGCAGC 1471
Qy 475 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValAlaGlySerPhe 494
Db 1472 ACCAAGCGCATCGACTACGCGGCAACAAAGCCCTTCCAAATCGTCCGTCGAGCTAGC 1531
Qy 495 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 514
Db 1532 GAC-----GACTATCCGACC-----CTGGCTCTGTCTCGAATCTCGGATCGACT 1576
Qy 515 TTPPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 534
Db 1577 TGTATATGACAGACTACGCCCGGTCAACAGCACCGGACCGGACGGGTGGTCTTTCCGCC 1636
Qy 535 AspGlySerArgPheValTTPAlaProGlyAspProGlyGlnProValValTyrAlaVal 554
Db 1637 GATGGTACACGGTGTCTTCATGTCCAGCACCTCGGCGGTGGTGTGCAAGTCGCAG 1696
Qy 555 GlyPheGlyAsnSerTTPAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 574
Db 1697 GGC-----ACCTTACTGGGTTCCTCTCTCCCGTCCGCGGGCGGTGATCGCTCC 1747
Qy 575 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 594
Db 1748 GACATCCGACCAACACGGTCTTCTATGTGGTGGATCCCGGGCGGCAATCTAGTGTGAG 1807
Qy 595 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 614
Db 1808 AACACAGCGACTCTCTTCCACCAAGACCGTGAGC---CTGGGTCCAGCACGACCGTCAAC 1864
Qy 615 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTTPLeuAlaAlaSerSerGly 634
Db 1865 GCCATC---CGCGCGCATCCCTCCATCGCGCGGATGCTCTGGCGGTGACCGCACAAAGGC 1921
Qy 635 LeuTyrHisSerThrAsnGlyGlySerSerTTPSerAlaIle---ThrGlyValSerSer 653
Db 1922 CTCTGGCATTCACGATTTACGGTAGTACCTTCCAGCAGATCCGCGAGTCCGCGGTACGGCG 1981
Qy 654 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 673
Db 1982 GGTGAGCTTGGGTGGCAAGCGCTCTCCACGGGACGCTAGTCTGATCTACCGG 2041
Qy 674 ValGlyThrIleGlyGlyValThrGlyValaryrArgSerAspCysGlyThrThrTTP 693
Db 2042 TTCTTCACCATCGAGTGGCGGGGTGTGTCAAGAGCGAGGACGAGGACCAACTGG 2101
Qy 694 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTTPGlyGlnAla-----IleThr 711
Db 2102 CAGTGTATCTCGGATGCTGCGACGATTCGGGTTCG---GGGTCCGCAACGCTGTGTAAC 2158
Qy 712 GlyAspHisAlaAsnLeuArgVal-TyrIleGlyThrAsnGlyArgGlyIleValTyr 731
Db 2159 GGGGACCTGAGACCTTACGGCGCGGTGTTCGTGGGCAACGACGCGCGGGCATTTCTTA 2218
Qy 731 rGlyAspIleGlyGlyAlaProSerGly 740
Db 2219 CGGCAATCCAGCGGGGACCTGCGCGGC 2246
```


RESULT 10
AY040839 3959 bp DNA linear PLN 01-APR-2002
LOCUS Aspergillus niger endoglucanase C (eglC) gene, complete cds.
DEFINITION Aspergillus niger endoglucanase C (eglC) gene, complete cds.
ACCESSION AY040839
VERSION AY040839.1 GI:19879405
KEYWORDS
SOURCE
ORGANISM Aspergillus niger
Aspergillus niger
Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 3959)
Haeper,A.A., Dekkers,E., van Mil,M., van de Vondervoort,P.J. and de
Graaff,L.H.
TITLE EglC, a new endoglucanase from Aspergillus niger with major
activity towards xyloglucan
JOURNAL Appl. Environ. Microbiol. 68 (4), 1556-1560 (2002)
MEDLINE 21914059
PUBMED 11916668
2 (bases 1 to 3959)
Haeper,A.A., Dekkers,E. and de Graaff,L.H.
REFERENCE Direct Submission
AUTHORS Submitted (18-JUN-2001) Molecular Genetics of Industrial
JOURNAL Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen
6703HA, The Netherlands

FEATURES
source
1..3959 Location/Qualifiers
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/db_xref="taxon:5061"
/gene="eglC"
/join(<1038..1137,1185..1329,1373..1571,1618..1858,
1908..1961,2007..>3841)
/gene="eglC"
/product="endoglucanase C"
/join(1038..1137,1185..1329,1373..1571,1618..1858,
1908..1961,2007..>3841)
/gene="eglC"
/codon_start=1
/product="endoglucanase C"
/protein_id="AAK7727.1"
/db_xref="GI:19879406"
/translation="MRRTFAALVAGYLLDSVHAASQAYTWKVVVTGGGGFTPGIVP
NPSAKGVAYRTDGGAYRLNSDDTWTFELMDWANNNSWHWDGIDAIATDPDTRVYV
YFGMTNDWDNDGSLRSDGTWETKLPFKVGNMPCRGVGERLAVDPNDSIL
YFGARSGLWKSIDYGTNSNVAFTKGTFFQDSSSTYSDPVGIAMVTFDSTGS
SGSPTRIFVGVVDTGESVFSVSDAGETWVSGEPMYGFLPHKGLSPSHTLYISY
SNGAGPYDGTNGTVHKNTSGVWTDISPTMTDTYGYGLAVDLOVPQFTVMVAALN
CWMPDELWRSTDGGSWSPIMWNGYPSINYYSDIENAPWLODSTDEPPVRVG
WVVEALIDPDSRHLVGTGETIYGGHDLQWDSHNVTIESLAVGIERNAVGLIT
PPGPAALLAVDDGGFVHTLUTTAPSYIHTPYSTNGIDIDYAGNKPANIVRSGSD
SDPTIALSSFGESWADYAAASSSTATQVALSDADATILLMNSDGYRNSATLSA
VSLPSGAVIASDKANNFYFASGSFYLSDTAATFTVTTLGSSSTANAIAPQS
LAGDVWTDGLHQTINWYKSFTOIGSGCTGEGWFGFKPSDGDYVPLVGFFTVDG
VTGLFKEDQGVNQLISDAHEGFGSASNVNGDLONYGRVFGTNGRGIFYGDPSG
TLPSATATASASSTAVKSTSTSTSKVGSSTVSSSTATTITITSSIKSTLTITTKS
SSSTSTSTSTATGASAYQCGGGSGFTGTPCGSGWCTCTYENEYYSQCKSPIGIATDR
G"

Alignment Scores:
Pred. No.: 5,646-74 Length: 3959
Score: 1620.00 Matches: 352
Percent Similarity: 55.57% Conservative: 107
Best Local Similarity: 42.62% Mismatches: 255
Query Match: 40.14% Indels: 113
Dbs: 8 Gaps: 20

US-09-917-376-3 (1-740) x AY040839 (1-3959)

QY 1 AlathrThrGlnProTyThrTrpSerAnValAlaile-----13

Db 1095 GCCCATCCAGGCATATACCTGGAAAGATGTGGTCAC-AGGGGGGTATGACATATCTTGC 1153
QY 14 -----GlyGlyGlyGlyPheValaspGlyIleVal 23
Db 1154 AATATGAGGATTGACAAAGCTAATGTATCAGCGCGGCGAGGCTTTCAGCGCGGTATCGTC 1213
QY 24 PheAsnGluGlyAlaProGlyIleLeuTyThrValArgThrAspIleGlyGlyMetTyArg 43
1214 TTCAACCTTTCAGCTAAAGGTGTGCGATATGTCGCACCGACATCGCGCGGTGATATCGG 1273
QY 44 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrp 63
1274 TTGAAC---TCGACGACATACATGACCCCGTGTGATGACTGGGCCCAACAACTCTAACTCG 1330
QY 64 -----GlyTyThr-Asn-----GlyValVa 69
1331 TAGTAAAGCTTCTTACAGATGCTAGCTCAGCTGCTCCAGGCACGATTTGGGGCATGA 1390
QY 69 IserIleAlaAlaAspProIleAsnThrAsnIysValTrpAlaAlaValGlyMetTyThr 89
1391 TGCATCGCGACCGACCGCGTGCATACCGATCGGTATATGTTGCAAGTTGGCATGTACAC 1450
QY 89 rAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTr 109
1451 CAACGACTGGGATCCCAATGACGGCTCCATCTCTCGGTTCGACCGACGAGCGGCATACCTG 1510
QY 109 pGlnIleThrProLeuProPheLeuGlyAsnMetProGlyArgGlyMetGlyGly 129
1511 GGAAGACAGAAAGCTTCCCTTCAAGTTGGAGGAATATGCCCGCTCGGGGTGTGGCGA 1570
QY 129 u-----ArgLeuAlaVala 134
1571 GGTAAAGTCTATATCTGGGTTTTGCTAGCATACACTGACCGCTGCGAGCTGTGGTGG 1630
QY 134 spProAsnAsnAspAsnIleLeuTyThrPheGlyAlaProSerGlyIysGlyLeuTrpArgS 154
1631 ATCCGAATGACAAACAGCATCTTGTACTTTGGTCCCGGAGTGGCAATGGCTTGTGGAAA 1690
QY 154 erThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrl 174
1691 GCACGAGCTACGCGGAGACATGGTCCATGTGACTGCTTCAAGTGGACCGGCACATTATT 1750
QY 174 leAlaAsnProThrAspThrThrGlyTyThrGlnSerAspIleGlnGlyValValTrpVala 194
1751 TCCAAGATTCAGCTCGACC-----TACACGCTGATCCGTTGCGGAATTCCTCGGTGCA 1804
QY 194 laPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheVal----- 211
1805 CCTTCGACTCGACCTCAGGTCAGTGGCTCTCCGACTCCGCGCATCTTTGTCGGTATGT 1864
QY 212 -----GlyValAlaAspProAsn 217
1865 GAAAAATTCTAGATATCTCAGGACGCGCCACTGACCATTTACAGGTGTGGGACACTGGGA 1924
QY 218 AsnProValPheTrpSerArgAspGlyGly----- 227
1925 GAATCGCTTTTGTGTCGAAAGATGCTGGGAAACTTGTAAAGCTTCTCCACGATCGCAAT 1984
QY 228 -----AlaThrTrpGlnAlaValProGlyAlaProThr---Gly 239
1985 TAAACCCATCTAACCAACTTAGGACTTGG-----GTATCTGGAGAGCGCATGTACGCG 2038
QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyThrIleAlaThrSer 259
2039 TTCTCTCCCCCATAAAGGACATCTCTGTCGCTTCTGAGCACACACTCTACATTTCTTACTCC 2098
QY 260 AsnThrGlyGlyProTyThrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279
2099 AACGGCGCGGTCCATACATGACCAATGACCGCTCCACAGTACATACATCACCACG 2158
QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyThrPheGly 299
2159 GGCGTCTGACCGACATCATGCTCCACCTCCATGACAGACACC-----TACTATGGC 2209


```
QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
DB 2210 TATGCGGAGCTGGAGTCTCAAGTGCAGGAACTGTGATGGTAGCAGCTCAAC 2269
QY 320 SerTrpTrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339
DB 2270 TGCTGGTGGCTGATGAGCTGATCTGGCGAGCAGCGGATCTTGGGGGACCTGGTCACCC 2329
QY 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359
DB 2330 ATCTGGGCTTGAAGCGCTATCCAGCATCACTACTACTACTCTTATGACATCTCCAAC 2389
QY 360 GluProTrpLeuThrPheGlyValGlnProAsnPro---ProValProSerProLysLeu 378
DB 2390 GCCCATGGCTACAGACGACACATCTACAGCAAGAAATTTCCCGTC-----CGCGTC 2440
QY 379 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 398
DB 2441 GGCCTGGATGGTCGAAGCCCTCGCCATTGATCCCTTCGACTCTGACCATCGCTTTACGGC 2500
QY 399 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIle 418
DB 2501 ACTGGCGAGACCATCTAGCGCGCCATGACCTCCAAATTTGGCAGCGAACAACACGTA 2560
QY 419 HisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro 438
DB 2561 ACTATCGAATCTCTTCGCGTAGGCATCGAAGAAATGGCTGTTCTGGGCTCATCACTCC 2620
QY 439 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp 458
DB 2621 CCAGCGCGGCTGCGCTTCTCTCGGTCGCGAGATGGCGGCTTCTACCACTTCC 2680
QY 459 ValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerVal 478
DB 2681 CTGACACCGCCCATCACTACATACACACCCGACTTACAGTTCAACAACCGCATC 2740
QY 479 AspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSer 498
DB 2741 GACTACGCGCGCAACAAGCCCGCTAACATCGTCGCTCTGGATCCAGGAT----- 2791
QY 499 GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly 518
DB 2792 -----TCGACCCCTACTCTCTCTTATCTCCACTTTCGCGCAATCTTGATACCGCAG 2845
QY 519 SerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg 538
DB 2846 TAGCGTGCTTCTCTCTACTCGACCGCGCAAGTCGCCCTCTACGCCGATCGACACAC 2905
QY 539 PheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsn 558
DB 2906 ATCTCTCTCATGACAGCGCGGT-----GCCATCGTTCCGCAAACTCGGCT 2953
QY 559 SerTrpAlaAspSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 578
DB 2954 ACTCTTAGCGCAGTCTCTCTCTCGCCCTCTGGAGCTGTCTATCGCCCTCCGACAAAGTAA 3013
QY 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 598
DB 3014 AACACCTACTTCTACGAGCGCTCCGGCTCTTCTCTATCTTTCATCCGACACCGCGCG 3073
QY 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
DB 3074 ACCTTC---ACTGTACACACACCTCGCAGTAGACACAACCGCAACCGCATC---CGC 3127
QY 619 AlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer 638
DB 3128 GCGCAACCTCTCTCGCAGCGCATGTCTGGTCAGTACAGACACAGAGTCTCTTCCACTCA 3187
QY 639 ThrAsnGlyLysSerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnVal 657
DB 3188 ACCAATACGCGCAATCTCTTCCACCAAGATTGGAGCGGATGCACAGAGGGGTGGAGTTTC 3247
```

```
QY 658 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIle 677
DB 3248 GGATTCGGAAGCCCTCTTCGGACGCGACTACCGGCTCTATTTGGATTCTTTACCGTC 3307
QY 678 GlyGlyValThrGlyAlaTyrArgSerAspArgCysGlyThrTrpValLeuIleAsn 697
DB 3308 GATGCTGTTCACCGGCTCTTCAAAACCGAAGATCAGGGGTAAATGGCAGATAATTTCC 3367
QY 698 AspAspGlnHisGlnTyrGlyAsn---TrpGlyGlnAlaIleThrGlyAspHisAlaAsn 716
DB 3368 GACCGCGAGCACGGTTCGGGCTCTCGCGCAAAATGATGATCAATGGGGATCTTCAGAAC 3427
QY 717 LeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGly 736
DB 3428 TATGCGAGGCTGTTGTTGGGACAAATGGCGAGGAATCTTCTATGCTGAT----- 3478
QY 737 AlaProSerGly 740
DB 3479 ---CCGAGTGA 3487

RESULT 11
LOCUS AY281371
DEFINITION Hypocrea jecorina strain QM6a Cel74a (cel74a) mRNA, complete cds.
ACCESSION AY281371
VERSION AY281371.1 GI:31747159
KEYWORDS
SOURCE
ORGANISM
Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
1. (bases 1 to 2517)
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J.M., Yao, J. and Ward, M.
Transcriptional Regulation of Biomass-Degrading Enzymes in the
Filamentous Fungus Trichoderma reesei
J. Biol. Chem. (2003) In press
2. (bases 1 to 2517)
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., Mitchell, T.,
Olivares, H.A., Teunissen, P.J.M., Yao, J. and Ward, M.
Direct Submission
Submitted (22-APR-2003) Genencor Intl., 925 Page Mill Road, Palo
Alto, CA 94304, USA
FEATURES
Location/Qualifiers
1..2517
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/notes="anamorph: Trichoderma reesei"
gene
1..2517
/gene="cel74a"
CDS
1..2517
/genes="cel74a"
/notes="putative secreted endoglucanase"
/codon_start=1
/product="Cel74a"
/protein_id="AAP57752.1"
/db_xref="GI:31747160"
/translation="MKVSRVLALVLGAVIPAHAFSWKNVKLGGGGFVPGIILPHPKT
KGVAYARTDIGLYRLNADDSWTAVTDGADNAGWNWGDIDAVALDDQDDQKVAAVAG
MTNSDPNGNGLIIRSDRGATFTNLFPKVGNGMPGRGGERLAVDPANSTIIFYG
ARSGNGLKSTGCGVTFSKVSSFTATGTYIPDPSDNGNSDKGLMWVTPDSTSTT
GGATSRIFVGTADNITASYVSTNAGSTWSAVPGQKYPFKAKLQPKATLYLASLNS
DGTGPDGTGLSGVVRDIAGTWKQIIPVSGSDLYFGFGGLDLDLQKPTLVASLNS
LTKRLGWMLESLEIDPDSNHWLYGTGMTIFFGGHDLTNWDRHNVSIOSLADGLEEPS
VODLAPAGGSELLAAVGDNDGFTFASRNDLGTSPOTWATPTWATSTSVYDAGNSVK
SVYRVGNTAGTQOAVISSDGGATWSIDYAADTSMNGGTWAYSDGDTILWSTASSGVQ
RSQFGSFPASVSLPAGAVIASDKTNSVYFAGSGSTFYYSKDTGSSFTRGPKLGSAG
```


Db 304 CACAATTGG-----GGAGTGGATGCGCTTCTTGATCCCGAGGACCCCAATGTG 354
Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
Db 355 GTGTACTTTCGGCGCGGGCTTGTACACCAACTCTCTGGGATCTTACCCCGCGGCATCTAC 414
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGly 120
Db 415 AAGAGCAGGACAGGGTGCACCTGGACATCTTCAACGCTCAGCTTCAAGGTGCGTGT 474
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
Db 475 AACATGCTGCGCGCGCATGGCGAGCGCGCTGCTCGATCCCAAGAACTCCCAACATT 534
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
Db 535 GTGTATTTCGCGCGCAGCTCTGGAAACCGGCTCTGGAGAGCACCGGACCGCGCAAGACC 594
Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 595 TTCAAGAGGTGCACTCTCAAGCTGTGTGTAACCTACATCCCGTCCCGAGGACAGC 654
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValTrpValAlaPheAspLysSerSer 200
Db 655 ACCGGCTACACACGACGACATCAACGGCTGACCTTTGTGACTTTGTACTCGACTCCGCA 714
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAsp---ProAsnAsnPro 219
Db 715 ACCATTAATGTGTCACCTGGCGCATATTCGTAGGCAACCGCCGATACGTCACGCGCTCG 774
Qy 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239
Db 775 CTTTACGTTTCTGAGGATCCGCGCAGCACATGGAAACGGGTCAAGGGCGACCTTGGCAA 834
Qy 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259
Db 835 TACTTTCTCATAGTCAGGCTCGAGCCACCGACCGGCGCACTGTACATCTTTACGCC 894
Qy 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279
Db 895 GACGAAACGGGCGCTTACGATGTGCTCTGGCGTGTCTGGCGATACGCGCTCGCGACC 954
Qy 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
Db 955 GGCACGTGGACGACATCACTCTCTCCGATGACGACCTGACC-----TACGGA 1005
Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
Db 1006 TTCGGCGGCTCTCCGTGGACATGAAGACGAGGAACCTGTGTGTCGCGCAGTCTCAAC 1065
Qy 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339
Db 1066 TCTTGTGGCTGTATGCGCAGATCTTCGCGACGACCGCACTCTGGCAAGACCTGTGTCAAG 1125
Qy 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359
Db 1126 CTCGTGGCAGTGGGTAACTATCCGAGATGATCTACTATTACGGCATCTCGACTCCCAAG 1185
Qy 360 GluProTrpLeuThrPheGlyValGlnProAsn---ProProValProSerProLysLeu 378
Db 1186 GCGCCCTGG-----ATCGGCNAGAACTTCTCTCGATACCGACTCCAAAGTGGCTG 1233
Qy 379 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 398
Db 1234 GGCTGGATGATCGAGTCTCTCGAGATCAACCGGACTGACAGCGACCACTTCTCTGTACGGC 1293
Qy 399 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIle 418
Db 1294 ACCGACTGACCCCTCTACGGTGGCCAGCAGCTGACCACTGGGACCAAGCGCAACATC 1353
Qy 419 HisIleAlaProMetValLysGlyLeuGluThrAlaValAlaAsnAspLeuIleSerPro 438

Db 1354 ACCATCCAGTCCCTGGCGCAGCGCATCGAGGAGACCGGTGTTCAGGAGGTGGCTCGGCC 1413
Qy 439 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis----- 456
Db 1414 CCGCGCGCAGCAGCTCATGGTGCACGCGGCGACATCTCGGGTTCACCTTCGCATCG 1473
Qy 457 ---AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGly 475
Db 1474 GCGGCAGACCTTGGAACTTCAACAAAGGGGTCTGGGACAAACCCCAAGTTCACCACTCA 1533
Qy 476 ThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAsp 495
Db 1534 GCCAGCTCGACTACGCGGCAACTCTCGCCGGGAAGGTGTGCGGTGGGCAAC---GAT 1590
Qy 496 ProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp 515
Db 1591 GCGCGGCTCAG-----ATGTCGCGAGCTCGACCGATGGCGGCTCATATGG 1638
Qy 516 PheGlnGlySerGluProGlyGlyValThr-----ThrGlyGlyThrValAlaAlaSer 533
Db 1639 -----AACGTGTACCCCGCGCGCACACAAAGAGCGGGTGCCTGTCTACTCT 1692
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTrAla 553
Db 1693 GCCGATGTGACACGATCGTGTGTGTCACCGGACCGGTGGC-----GTGTACCGC 1743
Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleAsz 573
Db 1744 TCCGTGAACACAGGCGAGTTTCTCCGCCGTGTCTCTGTGCAGAGCGGAGCTGTCTGGCC 1803
Qy 574 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 593
Db 1804 TCGCAACAGCGCAACACAGTACTTCTACGCGCATCCGCGACACTTTTCTACGTCTCC 1863
Qy 594 ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaVal 613
Db 1864 GCTGACGCGCGGTGACCTTT-----GCGCGCGCGGTCTCGCTCAGCGCGCCACG 1914
Qy 614 GlyValMet---PheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaAlaSer 632
Db 1915 TCGATCCGCGCATCGCGCCACCGACGCGCTTGGCAAGCTTATGTGTGCGACCAAC 1974
Qy 633 SerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThr---GlyVal 651
Db 1975 GCAGCGCTTTCACGACGACGACTTTCGGCAGACCTTTACCAAGACTTCCGCGCGCTG 2034
Qy 652 SerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaVal 671
Db 2035 AGCGACAGTACACAGATCGCGCTTGGCTGCTCT---GGTTCCAACCTGG---AACCTG 2088
Qy 672 PheValValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThr 691
Db 2089 TACGCTTTGCGACGCGCGCAACCGCGCAGGCTGTACGGCAGCGCGGCGGCGCTCG 2148
Qy 692 ThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGly-----AsnTrpGlyGlnAla 709
Db 2149 TCCTGGACTGATCTCCAGGGCGACAGCAGGGCTTCGGCATCATCAAC---GGATCGAG 2205
Qy 710 IleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIle 729
Db 2206 CTCGCTGGCAGTGGGAACGTGCGCAACAAGGTATACGTGGGACCAACCGCGCGCGCTC 2265
Qy 730 ValTyrGlyAsp-----IleGlyGlyAlaProSerGly 740
Db 2266 TTCAGGCGCAGGTTCCATCTCCGAGGCGCGCGCAGCAGCCGCTGTGT 2316
RESULT 13
AP005031/c 299550 bp DNA linear BCT 10-MAY-2003
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section
DEFINITION 11/30.
ACCESSION AP005031 BA000030
VERSION AP005031.1 GI:29606108

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
Gene
CDS

Streptomyces avermitilis MA-4680
Streptomyces avermitilis MA-4680
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.

1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M.
Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
21477403
11572948

2 Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,
Shiba, T., Sakaki, Y., Hattori, M., and Omura, S.
Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermitilis
Nat. Biotechnol. 21 (5), 526-531 (2003)
22608306
12692562

3 (bases 1 to 299550)
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kushida, N., Director-General of Biotechnology Center, Shiba, T.,
Sakaki, Y., and Hattori, M.
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail: bio@nite.go.jp, URL: <http://www.bio.nite.go.jp/>,
Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
Tel: 81-3-3481-1933, Fax: 81-3-3481-8424
Ishikawa (*2), Akiharu Hanamoto (*3), Chigusa Takahashi (*3), Jun
Shinose (*3), Hiroshi Horikawa (*4), Hidekazu Nakazawa (*4), Tomomi
Osonoe (*4), Norihiro Kushida (*4), Hisashi Kikuchi (*4), Tadayoshi
Shiba (*5), Yoshiyuki Sakaki (*6, *7), Masahira Hattori (*1, *7)
and Satoshi Omura (*1, *3).
Final finishing process and all annotation were done by H. Ikeda
and J. Ishikawa.

*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
<http://avermitilis.is.kitasato-u.ac.jp>.
Location/Qualifiers
1. .299550
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227882"
/note="This strain is also named as strain: ATCC 31367,
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
avermeccininus"
complement (42. .2936)
/gene="nrd"
complement (42. .2936)
/gene="nrd"
complement (42. .2936)
/note="SAV2461"
/codon_start=1
/transl_table=11
/product="putative vitamin B12-dependent ribonucleotide
reductase"
/protein_id="BAC70172.1"
/db_xref="GI:29606109"
/translation="MTETASGPARSGRAGTKAKGLRIERIHTTTCGVHPHYDEVEWARR
DYVMTNRDQSVNFSQGVFPDFWVSNVAINVTISKYFRGAVGTQFQREVSRLQIDRI

VKTRKAGEDYKYFASPADAEIFEHLAYALLHQIFSNFVWNVGTQPQVSAFC
ILAVDDSMESILDWYKBERGMIFKGGSGAGLNLRISSKXELSSGNGAGSPVSMRGA
DASACTIKSGDATRAAKWVILDVDPDIEDFETKVKKEEKIRALRDAGFDMDLGGD
DITTSVOYONANNSVNDTEFMKAVEGGKFGLTSMRTGVTIENLDAKSLFRKMAEAAW
ACADPGIYDDTINHHWTCPSGRINGNCPSEYMHLDNNTSCNLSLNLWKLKDDGK
GRQSEVERFAKVVVELVITAMDISICFADFPTQKIGENTRAFRLQIGIYANLALLMA
THGAYVSDGSLACAGAITSLMTGTSYKRSALAAVVPDYGVARNAPQHVNMKQSD
ANGAVRVDDLDTPIAAATAEAWQVHLKGNKGRFNAQSVIAPTGTIGLMSCDTT
GLEPDLAVKFKLGGGSMOIVNGTVFQALRRGLGQEIQEIAI VETIANGVNDIAP
GLKHEHYEFDGCMGERSISAMGHVMMAAIOPIWISGLSKTWNLPETADEVVEVY
FEAWKMGKALAIYRDNCKVGQPLSAKKETEKAEVTAKEATIRAEKVEKVPVR
KRLPKRPGITTSFTVGAEGYVTANSYDDGLGVEVFLKMSKQGSGLAGMDAFSTAV
SVGLQYGVPLTTSYKFTNMFEPATMDTDPDPRMAQSIVDYIFRLALDPLFETRS
ALGIHSAERQHELTGSEYSPSDVDMVEGLAQAPRAQELKAVATPKAEVAAVPA
PQQAHTSAELVEMOIGIOADAPLCFCGCTKMRAGSCYICEGCGSTSGCS"
complement (3086. .3649)
/note="SAV2462"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="BAC70173.1"
/db_xref="GI:29606110"
/translation="MHCPCPHDPSRVDSRTDDGTSIRRRQCDCSRRTVETC
SLMVKRSQVTEPFTKTVNGVRKACQGPVTEDALAQQRVEBAVATGSAELTT
HDVGLAILGPLRELDVAYLRPASVYTRAFPSLSDLEFATAELREQRERPAVDDHE
DAGAERQCTDRGSGGTVEVPATVAD"
4168. .4947
/gene="lexA"
4168. .4947
/gene="lexA"
/note="SAV2463"
/codon_start=1
/transl_table=11
/product="putative SOS regulatory protein LexA"
/protein_id="BAC70174.1"
/db_xref="GI:29606111"
/translation="MTTADSATITADRSQGRLEPVMAMNEATSHGPKRSIPGRPP
GIRASSGLTDRQRVIEVRUSVQRPYPSMREITGOAVGLSSSTSAVHQLMALERK
GFLRDPHPRAYEVRGSDQSSVQPTDTAGKPAASYVPLVGRRIAGGPILAESEVD
VPLPRQLVGDGELFVLKVGDSMIEAICDGDWTVTRQPVAEANGDIVAAMLDEAT
VKRFKREGHVWLLPHNSAYQIPGDEATILGKVVAVLRV"
complement (5110. .7143)
/gene="ding"
complement (5110. .7143)
/gene="ding"
/note="SAV2464"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent helicase"
/protein_id="BAC70175.1"
/db_xref="GI:29606112"
/translation="MTKPSLPHELLHAATAVAGTGPQVTMAEVAEADDSHLLV
AAGTKSGSLYLPALAHGERVIVATATLALQQLVERDLPTVIALHPLRRERPF
MLKGRSNYLCLHRLHEGVPODBEGLFQFEAAAPTSLGQDLRLRDSDETETGD
RDLTPGVSDRAWAQVSVSRCELGATKCAYGAECAEPMRSERAKLAERVVTHALLA
IDAIEGAPVPOHEVLIVDEAHLISRVTVGVAIGELTPGQVNRVRAAKLVNKEAD
QICTAAGFERMELAPGLREEIPDLGLVLMALRDAAITVISAIGATDKSQVDED
AVRKQALASVESVHDAERITNGSEWDVVMYERHDFRGLASRVPMSPVSGLLREKLF
DRSVLTATLKLGGDFNGVGSGLAPGEGTQGDLPQWKGVDVYSPFKQGLIY
AKHLSPARDGDRAMDDELTELIQAAGRTLGLFSMRQAALAAELRSRIPFPFIL
LQGEETLGLIKNPAADPTCLFGLTSLMQGVDPGSPSCOLVMDKIPFPDPDPLMS
ARQKAVEDAGGGMFAVAAHALLMQAGGRVLRASGDRGVAVLQRLATAGVSY
LKASLPDFWYTTDRNQVRSLTADEAAKKTETVTEVAEVAETEEAAKESGA"
complement (7376. .8509)
/note="SAV2465"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC70176.1"
/db_xref="GI:29606113"
/translation="MPPTASTDAEPATSDAPRIAVSARSAADSEDTLELRPDELV
ALPFAEAAAGGEAAGGAGVBERGTAGGERDVGCAEPAGRKDAGPBAQGTATA
AFSPARAGATASADSGCSDAVIDPAAFPAAASTVPSASRVPSAPSVTSAPSVTFAT

CDS	APAPPPSLGPTKALAGRAQVSVTRAADDLLGRVEDMGPTVTTPAGTFQALAPVRPERDL PLIGRWNDPAVAPFVWKLADGSETEQHLRAQLGQDGRSVFCLVGLGTPMSYWEIYR ADLSLARHYPARPHDTGTHILLIGVADRGRLGLSTLLRAVADLVLDRRPSCARVAE PDLENTSVSFAFLGAGFRFSAEVDLPDKRAALWDRDLRLDL" complement(8549..10609) /note="SAV2466" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="BAC70177.1" /db_xref="GI:29606114" /translation="MNATPTSDGSRSTHERAGHTDQLFTVQGPQAVFASAEQAAPAE TQAPLAPATQVFPVEVTPVQVASESVPPVGRARESERLRGTADLPDPEAQTA QAQAVENLLRCWRENGVLAPODGLRIPLLAGTALLVPVHWSPTGWHRFGLPYLA DAPQATPADAVTVVALLAREAGSRRAPRAHSAVDHCAVDHSDADLVGRVADSVRR AVFIDRRARADEPDLFLAABQALLLGHPLPTPKSREGISEARLYSPELRGSFA LHWLAVPVLATASWTERGRPVPAEQTLARLAGDGLPMPDGFALPLHPWQIREIR HRPAAALLDAGLRDGLPHGPWHPTSSVTRHTGAPAMKLKLSGLGRIITNSRREN RKEHLRGVEVHRLRSLGLEQQAHPGFQIVRPDPAVLAVDDQHGNAVPGDVMIRHN PFAPTDVSCVAGLSPRPMPQTEQPRVMSRLAEIIRLAGRTGRPGAVATEWF LRYLEVQVPRVLMDSAGIALEAHQNTLLLDPEGWPTGGRVDRNOGYVFRYSRA ELEARLPGIHSHTFVSDEVTDERFAYLGINNVGLIGLIGAFGSHLAEQQLLAAPR RFLADVATGPARLTPLETRLLDSPVLRCKANLTLRLHGLDELVGPDVTQSVYVTIAN PLHS" complement(10775..12229) /note="SAV2467" /codon_start=1 /transl_table=1 /product="putative class III aminotransferase" /protein_id="BAC70178.1" /db_xref="GI:29606115" /translation="MAVTKSAVASLKREAGHAHAHEGILRRQSAESAARTYARALP IVPYRAGLTLEGADRGVLDCLSGAGTLAGHNPVVLVLAIRKVLDSGAPLHVLDLA TPVDKATTEFLRTLPLAGLADRARVQFCGPAGTDAVEAAFKLVRATRGSLIAFTGA YHMTAGALEASGATDGRVARLPYPQFCGPGFISGGERGAELGARWTESLDDVKS VLPAGMLLEPVQEGGVIPAPDAWLRMRATITADRSIPLIADSVQTVGRTGRFWAV EHSQVPVDMVAGSLGSPPLAVVYRDDLDVWQPGAHAGTFRGNQLAMAAAGTATLA YVRHRLAERACTLGLARMLDLRLSLATEFCPCVGEVRGRLMIGVELVRAEDRAVSP GSDPFASLDITAGAPRAPELAAVQRECLRGILVELGGRHASVWRLPLPLTISDE QATAVLDRLADAEVAVARGAERSRGFORSERAG" 12761..13993 /note="SAV2468" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="BAC70179.1" /db_xref="GI:29606116" /translation="MRPRLPLFAARRGRSARRRTPVLGAVGLATATATACDSGD DTKASSESSPASAGDDKIKIPDDIRDKLKERGIDIDKWKNGAWKNWRDMLREAN	Db	161381	TACCGTGGCGCAACGCCGTATCGGGGGCACCGCGTTCGTACCGCGCGTCTCTCCAC	161322
	7.58e-61	Length:	398550		
	1405.00	Matches:	308		
	Percent Similarity:	Conservative:	104		
	Best local Similarity:	Mismatches:	257		
	Query Match:	Indels:	86		
	DB:	Gaps:	21		
	US-09-917-376-3 (1-740) x AP005031 (1-299550)				
Qy	6	TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn	25		
Db	161381	TACCGTGGCGCAACGCCGTATCGGGGGCACCGCGTTCGTACCGCGCGTCTCTCCAC	161322		
Qy	26	GlucGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp	45		
Db	161321	CCCTCCGTACGCGGTCTCGCTACCGCCGACCGACATCGCGCGCGCCCTACCGCTGGGAC	161262		
Qy	46	AlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	65		
Db	161261	GACCGCGCGCCCGCTCGACCCCGCTCATGCACACCTCGCGCTGGGACGACTGGAACCTC	161202		
Qy	66	AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal	85		

Alignment Scores:

Pred. No.:	7,58e-61	Length:	299550
Score:	1405.00	Matches:	308
Percent Similarity:	54.57%	Conservative:	104
Best Local Similarity:	40.79%	Mismatches:	257
Query Match:	34.81%	Indels:	86
DB:	1	Gaps:	21
US-09-917-376-3 (1-740) x AP005031 (1-299550)			
Qy	6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn	25	
Db	161381 TACCGCTGGCGCAACCGCGTACCGCGGGGACCGGCTTCGTACCGCGGTCTCTTCCAC	161322	
Qy	26 GluGlyAlaProGlyIleLeuTyrValArgThrAspileGlyGlyMetTyrArgTrpAsp	45	
Db	161321 CCCTCCGTACCGGCTCGCTACCGCGGACGACATCGCGGCGCGCTACCGCTGGGAC	161262	
Qy	46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	65	
Db	161261 GACCGCGCGCGCGCTGACCCGCTCATCGACCACTCGCGCTGGGACGACTGGAACCTC	161202	
Qy	66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal	85	

GWGAVAVDPQHPQVLLASTFRRTTRPRDEVRFGSGDGRSWVPLLLAQAQFDHSAAPWTAA
 ATPHWIGALAIIDPPSDNSHAFVTVGYGYNWASNLQTFSEQQPLOWFOQDGLQVVRSET
 LLSPMAGHLLSALGDIIDGRHDALDTAQLYLGPRTNGBSIDAGASQVWVPOQVVRST
 VRDRNEIRALYSQDGGTHWAFAFSEPPRQAGATJALADASQVWVPOQVVRST
 GDFGRWQVGLPDTTAVADRVDQKWTAAADRVSGLRYESSDGAASFRDTGQGVGS
 PARDERARPLRPDPFVAGVYVYASPTLGVNWRQQLRTLSKDSKARSGLIGIKALRA
 GAPPALYLAGRAGVVDGI FRSDDRGAWRRINDDAHRFGKPYSVTGDPRTAGRYVFAT
 GGRGIFVGDPR
 2431. 4347
 /gene="XAC1771"
 /gene="XAC1771"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="sialic acid-specific 9-O-acetyl esterase"
 /protein_id="AA036635.1"
 /db_xref="GI:21107969"
 /translation="MTASFVRRCGLLGLVLPSCAWVPTPLLLADGAVLQRPQMP
 VMGSSPNAATAVSFDCKRATVKADATQGWKVRLPAAAGGYPVLRVQDGGELQVRD
 VLVDVWLAGQSNWEPHQAQSDGQPAVAANDQRLQFKVFKSVMQPOARLTGGE
 WKAATPNAAGFTAVGFFAKELRASTGVPVIGVNSTWGGSAIEAMWDAASLGNADQ
 NQGAIEAIKORDAAQAATGKRIARVKEDEMPQWREAFDDSDMSIPVTKWESS
 GYDGMGIATWRTTITLSAAEAKGITLGVQCIIDSDTYYVNGQQVDSITEKQWNLPRV
 YQVPAALKAGWNHIAVRVEDLSGGGMHGPDEQRFVQTSAGAKRALDGMKFRPAARV
 VSLTDKNQLPTLYNOMIHPLOPPVKGVTIYQGETNATDTGAKVRYEQAPAIROW
 RAERGDKITPLWQLQANFKAGGDKGELSPWALLRESQSKTLALPAGQAVIIDIQNP
 TDHPTNKRQVGRHLAALAHVAYGETLVASVPFKASFDGSKAVLGFIDQGSALQV
 RGGVAVQGFRIAGADQRFHPATAQIDGDRVTVRSDAVAVAVRYGWSNPENPDANLIN
 RDALPVSPFRTDTW"
 4344. 6101
 /gene="XAC1772"
 /gene="XAC1772"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA036636.1"
 /db_xref="GI:21107970"
 /translation="MTRSLMSMHFVARVIASSPARRRLRPLLSALALAAACQSGE
 QDQKAPAAPIAGSAATTPATAAPAAQTPMPQFVSKDGKHALMVDGAPFLILGAQ
 VNNSNYPGIMDKVWPAMQVLPNTVQVPIAWEQVEPEGREFSFDVTLNQAREHK
 VRLVLLMFWATWKNQGPAPHWKTDNPFRRVTRREGSGISLSPHAQATLDADKA
 FVAFMQLKKYDPPQHTVIMIQIPENEAGTYSVRDPSMAQKLPDQVDELIRKNNQ
 PGTWAQFQADADIFWASIGRYIDQIAQAKAEPLPMTVNAALRGFPNPGQGY
 ASGGPTDNLDVWKAAGHIDLLAPDIYMEYRYMTVTLERYAPDNALFVAETGNRP
 EYARYLPTLGHNGIWSAGFIDYTRYSNYPLGAKHVNTEETLAPALGFOAVSGMRP
 FAKAASEGKLHGTAEPCQAVQBELKLNDRWSATITYGVQFWFKGEPGPNPEPIGAL
 IAEUSPDEFVLTVGHARVTLHPASDATANMIYDRVEGVYEGNDWKFORNWNQDQTDY
 GWNFSVPQLLKILATYK"
 5999. 9109
 /gene="xyls"
 /note="synonym: XAC1773"
 5999. 9109
 /gene="xyls"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="alpha-xylosidase"
 /protein_id="AA036637.1"
 /db_xref="GI:21107971"
 /translation="MEIPAOLEWRPDRLRROFFORAAAQOAGNLQISAPATPRRR
 RDGTAPVPCAPAPAPRHSKVMOTTIERSAGAMPANRRTSLAQCLALSIALLANA
 AHAEVRKAAGVTVFSAKQAPVRLVVDVADGIIRVSADPDGDFARGSPSLMRVPVQ
 DTAFCVAEQSGSVLQTKTASISTVDGHSFADANGKPVLSVEFARGSPFAPLNVEG
 KOYLSVRQSPDDEALYFGFQHQQNMGNRVELQNNIDMAVYLSNRYNGY
 LWDNNISITRLGDPRLQPKTLTLYDAKAGKAGALTARYANGHILRRSESVNYQY
 LSDLTQPKKATTKDKNRSMQVWTEGETEALTGGHTFSLYSSEYAKLYVDGKLVDVR
 WRQNWPNWNEFKLEPGQRHTVYKVEWDLIDPSIALLHRDPLPAAEAKDLSLWSEA

GQMIDYFVSADSDQAQVAGYRALTGKSTMLPKWAYGFQWQSRERYKSQBELVGAAVEY
 RTKSLDNI VLDMSPENAWGSHDFDQHFDPDPDMVKAVHDMHAQIMISWPKFY
 PPKYKELDAAGFWPKRNEVGEIDMTGKGYKNSFYDPYSEKAQAIYWRQINKEKLS
 KGFDAWMDADEPQVHNSLDIAERKARTPNALGSSSTEVFNSYPLPHTHGVYVGDRAA
 DDKVFPILSRKGYAGTQRNAVAVMSGDI VSRWDDMRDOI SGGVNNALSGLPNWTFDIG
 GFVAKRYEDQDPAHLPEWRELNTRTFQGFAPVPLFRSHGQFPYREIWNIAPEGTFFY
 ESMAYNRLRSALLPYIYSLAGDTTYORDGVIRMGMMDFPNDPKVDINDQYDFGFAP
 LVAPYTRFGATSRQVLPAGSWSLEFATGKRYEGGCSIEAAAPIERMPILFVRGASIVP
 TGPVOEYVDQVADAPLTVVVTGADQFSLYEDDGGKGYEGKFSRILPLVWNOAKGE
 LSIKRGESWTGMQAKRTINRVFVDGPRDDAGALAPKTDASIOYDGKPVSVLQKTIAS
 GKARRR"
 9201. 11573
 /gene="XAC1774"
 9201. 11573
 /gene="XAC1774"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA036638.1"
 /db_xref="GI:21107972"
 /translation="MLTRRELCKATGAIAVLGAVPAAQAQATQASKTALPVAAAEA
 LQMTREPANWEALPVNGRKLGAWVGGAIAHERLQLNEDTLIYAGGPDSTSPDALA
 ALPQVRAILFAGRYAEAEKLAADAKLLSRPLKQMPYQPLDGLLLDFDRAIGISDYRQL
 DLDTAVATTFRSGAVHRRVFCQAQCI VVRLSCDRPGGILSRVIGDSPTQGEVT
 AEPGGLLPSGRNSFAGIEGRLFALRVLPOVSGGKLQSVDRDLRIDAADDEVVLLLSA
 ATSYORFADVDGDPALTAARLKAALDFFALLRAHLADHORLPRRVAIDIGSSRAV
 QLTDERVQRFAGNDPALAALYHQYGRYILLICSSRPCTGPANLQIWNLMQPPWES
 QLTININEMNTWSEANALHECVLEPLEALDLAGTGTHTATAIYDAGPWVHNHND
 LWRQAGPTDGAQSLMPNGVWVLLQQLMDRWDYDRAYLSKVYPLFGAAEFVATL
 MRDPQTAGMVTNPSMSPENQHPFGAAVCAQSGMDAQLLDLFAQCIAMSKLLIGIDQL
 AQLAALREQLPPNRIGKAGLOEWOQWDMQAPENHHRVSHLYALHPSSQIINLRDT
 PELAAARSLEIGDNATGMIGIWRNLNLAELADGEHAYRILQLLISPERTYPLNLF
 AHPPQIDGNFGGTAGITMELLOSWSGVFLPALPKWPRGSRVGRGLRVRGGSASVLE
 WEGGRLQARLHSDRGGRYQUSYAGQTLDELGAGRTQOVGLNNRLVTQ"
 11570. 13063
 /gene="XAC1775"
 11570. 13063
 /gene="XAC1775"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="D-xylosokinase"
 /protein_id="AA036639.1"
 /db_xref="GI:21107973"
 /translation="MSLYVGLDVGTSQVKLVAYDPQDRAVAVATIAAPMELISRDGTR
 EQQAQWIDGIVHCFQALDGEQARVRGIVSGQQHGFVPAADGSVTPAPVKLWCDS

Alignment Scores:
 Pred. No.: 6,748-50 Length: 14520
 Score: 1162.50 Matches: 268
 Percent Similarity: 51.07% Conservative: 115
 Best Local Similarity: 35.73% Mismatches: 306
 Query Match: 28.80% Indels: 61
 DB: 1 Gaps: 20

US-09-917-376-3 (1-740) x AE011809 (1-14520)

QY 2 ThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAspGly 21
 Db 213 ACGCCAGCCCGATATCAGTGGCAGCGTGGCCATCGGGGGGGCGGCTTTGTCCCGGT 272
 QY 22 IleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMet 41
 Db 273 GTGCTGTTTCATCCCGCAACGCGATCTTCCTATGCACGCCGATGTCGGTGGTGC 332
 QY 42 TyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsn 61
 Db 333 TATCGTGGGACCGCGAGCGCGAGTGGCTCGCGCTGACCGATTGGTTGGGCGCGAC 392
 QY 62 AsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal 81


```
Db      393  GATTGGAATCTGATGGGAATCGACGCTTTTCGGTTCGATCTTCGCCGACCCGGAACGCGTGG 452
Qy      82  TrpAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg 101
Db      453  TACCTGGCTGTGGCACCCTACATCAGCAACGCGCCGCAAT---GGCGCGGTGTGGCT 509
Qy      102  SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyValAsn 121
Db      510  TCGTTTCGATCCGGTAAGCAGTTGGAACGCCGCCGACCTGCCCTTCAGCTCGTGGCAAT 569
Qy      122  MetProGlyArgGlyMetGlyValGlyLeuAlaValAspProAsnAspAsnIleLeu 141
Db      570  CAATTGGCGCGCCCAATGGCAGCGCTGCGCTGACCCGCCACGACGCGCGGTGCTG 629
Qy      142  TyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 161
Db      630  TTGCTTGGCTCACGTGATGCG---GGCCTGTGGCGCAGCAGCATCGCGCGCGCACTGG 686
Qy      162  SerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 181
Db      687  GCGCGGTGGGAAGATTTCCTGCCGATGCATTTGGCTGGCGCCACGCGCGCAACCATGTC 746
Qy      182  GlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 201
Db      747  GGAAGCAACAGCGCGTG---GGCATCGCTTCGTGGTTCGACGCGCGCCAGCGCCAT 803
Qy      202  LeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe 221
Db      804  GCTGGCGCTGCGACACCAACGATCTATGTGGCGTTTCGACTGGCGCAACACGACCTGTAT 863
Qy      222  TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle 241
Db      864  GTCTCCGAAGATGGGGCGCACCTGTGTGGCGCTGCGCGCGGCAACCAAGCGCGCTGCGC 923
Qy      242  ProHisLysGlyValPheAspProValAsnHisValLeu----- 254
Db      924  CCC-----AGCCACATGGTTCGGTCAACACGCTGGCGGAGTGG 959
Qy      255  TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 274
Db      960  TACTTGAGTATGGGATCGCCCGCGCGGATCTGATGGCGGTGGCGCTGTGGAAAC 1019
Qy      275  PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 294
Db      1020  TACGACGCGACGCAAGCGCGCTGGCGCGAGATCAGCCGATTCACAAACACGACCACT--- 1076
Qy      295  AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 314
Db      1077  GGCATGTGTTTCGGCTGGCGCGCGGTGGCGGTGGATCCGCAACATCCGAGGTGCTGCTG 1136
Qy      315  ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 334
Db      1137  GCCAGCAGTTCCGCGCGCGCACGCGCGCGATGAGTGTTCGACGCGCGGATGGCGGA 1196
Qy      335  AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 354
Db      1197  GGCAGCTGGTTCGCTGTGGCGCTGGCGAGTTC----- 1232
Qy      355  LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 374
Db      1233  ---GATCACAGCGCGCGCGCTGG---ACCGCGCACGCAACACCCGAC 1274
Qy      375  SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394
Db      1275  -----TGATAGCC---GGCTGGCGATCGATCCGTTTCATAGCAACCAAC 1316
Qy      395  MetLeuTyrGlyThrGlyAlaThrLeuTyrThrAlaThrAsnAspLeuThrLysTrpAspSer 414
Db      1317  GCCAGTTTCGTCACCGCGCTACGGCATCGGGCTCGCGCAATCTCAGACGTTTTCAGAG 1376
Qy      415  GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434
```

```
Db      1377  CAGCAGCCGCTGCACTGTGTGGTTTCAGAGCCGTGGGCTGGAGAAACCGTGGCGCTGGAT 1436
Qy      435  LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454
Db      1437  CTGCTCAGCCGATGGCGCGCGGATCTGCTCAGCGCGCTGGCGGATATCGATGGCTTC 1496
Qy      455  ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 474
Db      1497  CGGCAT---GACCGCTGGACACCGCACAGTTCAGATATCTTGGTCCGCGCTGACCAAT 1553
Qy      475  GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 494
Db      1554  GCGCAAGATCATGATGGCGCGCGCCAGCGCGCGCTGTATTCGACGACGCGGACCGTGG 1613
Qy      495  AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 514
Db      1614  ---CCGCGATCTGCGCAACACAGATTCGCGCGCTGTATTCGACGACGCGGCGCACGCAC 1670
Qy      515  TrpPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 533
Db      1671  TGGGCTGCATTCGCCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1730
Qy      534  AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 553
Db      1731  GCGGATGCATCGCAGGTGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 1781
Qy      554  ValGly---PheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 572
Db      1782  ACTGCGCATTCGCCAAGCGTGGCAGCGCTGCAGGCGCTTGCCTGATCCGCGGTGGT 1841
Qy      573  ArgSerAspArgValAsnProLysThrPheTyrAlaLeuSer-----AsnGlyThrPhe 590
Db      1842  GTGGCGGACCGGTGCATCGCAGCGCTGTATTCGCGCGCATCGGTGGAGCGCGCGCTG 1901
Qy      591  TyrArgSerThrAspGlyGlyValThrPheGlnProValAla-----AlaGlyLeuPro 608
Db      1902  TACGAGAGCAGCATGGCGCGCGCGCTTTCGCGATACCGACAGCAGGTGGCAGCGCG 1961
Qy      609  SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlyValAspLeuTrp 628
Db      1962  GCGCGCGATGAA-----CGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2006
Qy      629  LeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSer----- 646
Db      2007  CGTGGCGCGGTGTGTATCTGGCCAGCCCGCGCTGGCGGTGATCGCTGGCAGACCGC 2066
Qy      647  -----AlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla 663
Db      2067  CAGTTGCGCACGCTATCCAAATCAGATGAAGCGCGCTCGCTGGGCATCGCGAAGCGCTA 2126
Qy      664  ProGlySerSerTyrProAlaValPheValGlyThrIleGlyGlyValThrGlyAla 683
Db      2127  CGTGGCGCGCGCGCGCGCGCTGTATCTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2186
Qy      684  TyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyr 703
Db      2187  TTTCGCTCCGACGACGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2246
Qy      704  GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGly 723
Db      2247  GGCNAA---CCGTATAGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2303
Qy      724  ThrAsnGlyArgGlyIleValTyrGlyAsp 733
Db      2304  ACCGCGCGCGGTGGCATCTTCTATGGCAT 2333
```

RESULT 15

AE001712
LOCUS
DEFINITION
Thermotoga maritima MSB8 section 24 of 136 of the complete genome.
AE001712
ACCESSION
AE001712.1
VERSION
GI:4980799
KEYWORDS

gene	3073..3942
CDS	/locus_tag="TM0302"
	3073..3942
	/locus_tag="TM0302"
	/note="similar to GB:Pyro.h percent identity: 57.09; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="oligopeptide ABC transporter, permease protein"
	/protein_id="AAD35390.1"
	/db_xref="GI:4980802"
	/translation="MKRILRLFFDMWEDLRFKYAVTVLTALVLLSLSPSPVDPYKMWYLPNTQPPSLEHILGTGKQDIFWLLTFSKNSLILATIAAFSRIIATIGMLSGYLGTTDRILMALTDITFWMPVFPFLLLLSLKNLYLSPLGLIIGVFGWAGDARVIRSMILSLREREPVTSKFSQMTFEEVFGDFVPLIPVISGGFPGISGMAIGAEIVLAILGTRVRIPTLGSNFMHINPQAMLLGWWVLTPIVAVLFTALYTLISLSINEYLDPRTEMORIGAVK"
gene	3945..4931
CDS	/locus_tag="TM0303"
	3945..4931
	/locus_tag="TM0303"
	/note="similar to GB:U08186 PID:473270 GB:AE000512 percent identity: 76.00; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="oligopeptide ABC transporter, ATP-binding protein"
	/protein_id="AAD35391.1"
	/db_xref="GI:4980803"
	/translation="MEKIAVAENLRAYFIKSGDKTTSIRAVDNVSLHFFENEVYGIAGESGCKSTLLKVLFEIKPPLKVDGEVVKLDGEEYIRNSMSLDQLKIRWKYVSYPQGSMSVLNPKRIERIFLDVIEHTSMREAFSRIESYFALGLPVRLVDSYHQVSGMRQRTTIALATVNLNPKVIFADBPPTALDVVQORGIQLLKTHKEQNTLIVITHMGVHAHLTRIGVYAGRLIEAEATHIEIKNPLHPYTRFLNSLPFRGDRSRKEGTPEGSPSLANLPSCGCFHPRCPAQICMEPEPDLIEVARNHKVACWKVGVER"
gene	4928..5878
CDS	/locus_tag="TM0304"
	4928..5878
	/locus_tag="TM0304"
	/note="similar to GB:Pyro.h percent identity: 65.51; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="oligopeptide ABC transporter, ATP-binding protein"
	/protein_id="AAD35392.1"
	/db_xref="GI:4980804"
	/translation="MRVSVRNLTKIFRVGSIFSLKKLVADINVSFEIGSGEIFSLAGESGCGTITAKILLGLEEPTNGEVLVDGILKDFVENEMHFLKKVQTVQONPFSFNPFRKVEDYLFKTVLRFKMAVNTKDALRLIEKLSVGLNFEIKRYSEFSGGQLQRISARALLTPSFVADPEVSMIDASURISIVNLFRDLKRGFGISILYITHDLSTAIYVA DKLAIMFRGSI VEMGEAQDILDTPHPTTKLLKLSIPVDPDEYENDENINLSSTEREEFPKAGCKPASRCVMSICRESVPPDFDVNGRVVVKWLK"
gene	5894..8017
CDS	/locus_tag="TM0305"
	5894..8017
	/locus_tag="TM0305"
	/note="similar to GP:3560020 percent identity: 55.11; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="endoglucanase, putative"
	/protein_id="AAD35393.1"
	/db_xref="GI:4980805"
	/translation="MLRSFLILFLAILGVFGATFEWKSVEINGCGFVPGIIFHPASPGLLYATQWAGYGAIIISDYGETWTIVNLKDYKVGNGEDNRAGERLQVDNPFNSVLFM GTTKYGLMKSDEFGKWKVDSFSTSTVFLVDFKSSEKSPTRI FVGCSEPKGI FVTEDGGTTNNVLNPNLDLIPURGIKHGILYITUSNAGPNAGTRGAVKMYVIADQK WYDVTMPKIDNPHWISDVKIDPDMNRAI FTTGTGYVMVTVYELKSKSEFGKPKWFIEN NKAPIRETVVIVLPPIGRPLLSAIADWGGFRHESLDTPPSSMYKELKWTSLGIATFAYQ RGI

NSKFARVHTYTPFLSYSDGGINWREIETVPEGITDGGRLSLAVNSDGGKTLVWSPA
 NHEVIVSSDKGSKWKAISVPPEFENYPPASDPNPSKEVIFDWNKGDFLISXKGGKS
 FMKAGPFDNDWWSLYSPVLAPDRBGIDWALQWNGLYRSKGGITPERRLGNVDI
 AYVIFGAPKPGDYDPAIYLNGWNGVYIFNSTDEGKTWNRINNDKHQFQWIIHMLG
 DWNFGRIFLGTGRGRIIVGEVKEE"
 8014..9363

gene

/locus_tag="TM0306"

CDS

8014..9363

/locus_tag="TM0306"

/notes="similar to GB:M29877 SP:P04066 PID:1335066

PID:178409 PID:182779 percent identity: 56.91; identified

by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="alpha-L-fucosidase, putative"

/protein_id="AAD35394.1"

/db_xref="GI:4980806"

/translations="M18KPKRYPDWESLREHTVPKPFDAKAKGFIHWMGIYSPGWA
 TPTEGLGKVPMDAEPFONPAEYVENSIRIKESPTWEYHVKTGYGENFEYKPADLPTA
 EKWDQEMADLFKKAGAKYVITPKKHDGCLNGTKYTDNSVVRGPKRDLVGLAKA
 VREAGLRFGVYSGGLDMRFTEPIRYPEDLSYIRPNTEYADYAKQVMELVDLIP
 DVLNDMGWPEKGEDUKYLFAYYINKHPEGSVNDRWGVPKMDFKTAHYHVNYPGDL
 KYWFEFTGLGFGYRNEGPEHMLSVBQLVYTLVDVWSKGNLLNVPKPGDGTIP
 DLKERLLGGEWLRYGDALYGTSVWERCACAKTEDETEIRFTRKCNRI FVIFIGIPT
 GEKIVIEDNLASGTVRHPITGERLSFNKVGKNLEIIVPKKLETDSDITLVLEAVEE"
 9360..10781

gene

/locus_tag="TM0307"

CDS

9360..10781

/locus_tag="TM0307"

/notes="similar to GB:L42023 SP:P44779 PID:1005445

PID:11220691 PID:1204865 percent identity: 56.91;

identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="L-fucose isomerase, putative"

/protein_id="AAD35395.1"

/db_xref="GI:4980807"

/translation="MKGKYRVGLISFSDGRBYVHRELEGIIVKVFEDIKLTLEBTEGV
 EVIRAKIIVKPSINAKREARLAREGEAVITFNVAWPFHTVLAKFPAGPFLIFS
 NINPOYPMWAMLSAAGALEODGKHYRMGMGNIKDEKVLKRVMAPIRAGSYPKLRGQ
 RYVGSGSMGWTAVPNVDLWKKIFGVDEHIDQFIVRESQLIPDRARAKRLWIE
 EKAKHYDKYLTPEKLEQIKSYHVAIVEEMELD FVGIKQGLELTFEFTVMDVT
 ETLNDYDMEGEHPIVCAEDSDGALTMQIFKLAKTLPVLFADVNRHYVEEDVLD
 LCNNGNHATYFAARSPDENKNKVFYPTQFPAGAAVKHAAPOKVTGLRLTRE
 DGRYRFTVPGFVDFGEEKNYEIAADSIQNNWPHAFUKMETPIDEFLLAKYSNNIHGV
 YGDYVERIKTFCEIASIDFVLMK"
 complement (10798..13092)

gene

/locus_tag="TM0308"

CDS

complement (10798..13092)

Alignment Scores:

Pred. No.:	7,72e-50	Length:	13278
Score:	1160.50	Matches:	264
Percent Similarity:	51.53%	Conservative:	123
Best Local Similarity:	35.15%	Mismatches:	275
Query Match:	28.75%	Indels:	89
DB:	1	Gaps:	26

US-09-917-376-3 (1-740) x AB001712 (1-13278)

Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
 Db 5954 TTTCAGTGGAAATCGGTGGAATCAACGAGGAGGATTTGTACAGGAATAATATTCAT 6013
 Qy 26 GluGlyAlaProGlyIleLeuTyrrValargThrAspIleGlyGlyMetTyrrArgTrpAsp 45
 Db 6014 CCTGCTAGTCCTGGGCTTTTATATGCGAAGACGGATGTTGGTGGGCTTTACAGGGTGGAT 6073
 Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyr 65
 Db 6074 GAGGAACAAGAGGTGGAAGCAATTTATGATTTTITTTGAGNAGAGATCAGTCCGATATAC 6133
 Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85

Db 6134 ATGGAGGTGTTGACGGCTTGCTCTCGATCTCTAGTCCGAAAGAGAAATCTACGCGATGACA 6193
 Qy 86 GlyMetTyrrThrAsnSerTrpAspProAsnAspGlyValAlaIleLeuArgSerSerAspGln 105
 Db 6194 GGAATAACATCAGTACAGATTTGG---GCAGGATATGGAGCCATTCTTATCTCGAAGATTAC 6250
 Qy 106 GlyAlaThrTrpGlnIleThrProLeuPro-----PheLysLeuGlyGlyAsnMet 122
 Db 6251 GGTGAAACGTGGACCAATTGTAAACCTTGCAAGATATGGGATAAAGTTGTTGGGAACGAA 6310
 Qy 123 ProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr 142
 Db 6311 GATGGAAGAAACCGCTGTTGAAAGATTGCAAGTAGATCCAAATTTTCAGTAGCGTCTTTT 6370
 Qy 143 PheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSer 162
 Db 6371 ATGGGTACA---ACGAAATATCGTCTGTGGAATCAGAGATTTTGGAAAAAATTTGAAA 6427
 Qy 163 GlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly 182
 Db 6428 AAAGTGGATAGTTTTCCCTCAACC-----AGTCTAACGTTTGTCTATTCGATGAGAAAAAGTGGAAAA 6493
 Qy 183 TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeu 202
 Db 6494 GGAAGTCTCTACTCTAGAAATTTTGTGGATGTTTCAGAACCCAGGGG---ATCTTTGTA 6550
 Qy 223 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro 242
 Db 6551 ACTGAAGATGGTGGAAACCACTGGAAATGTTCTTCCAAACCTCGCAACGATCTCATTTCCA 6610
 Qy 243 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 262
 Db 6611 TTGCGTGGAAAAATACAT-----GACGGGATCTTTACGTTACACATTTCCATGCTTTG 6664
 Qy 263 GlyProTyrrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp 282
 Db 6665 GCACCGAATGGTGTACGAGAGGGGCTGTTATGAAGTATGATGATGCTGATCAAAAAATGG 6724
 Qy 283 ThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGly 302
 Db 6725 TACGATGTAAACCAATGAAGGAGAT-----TTCCGATATTTGTTGA 6766
 Qy 303 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp 322
 Db 6767 ATAGATGTTCAAGAG-----AACGTGGTGTATGATATCCACATCGATAGTGGTAT 6817
 Qy 323 ProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp 342
 Db 6818 CCACACGATGAGATATTTATTTCCCTGAACGAGGAGAGAAACGTGGAGACCTCTTCTGGAA 6877
 Qy 343 TrpThrSerTyrrProAsnArgSerLeuArgTyrrValLeuAspIleSerAlaGluProTrp 362
 Db 6878 AAGGCCAAATTTT-----GATATCAACAAACGCGCCCTGG 6910
 Qy 363 LeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAsp 382
 Db 6911 ATAAAA-----GATTTAAATCTT-----CACTGGATATCT 6940
 Qy 383 GluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrrGlyThrGlyAlaThr 402
 Db 6941 GAC---GTCAAGATAGATCCATTTGACATGAATAGGCTATTTTTCACACAGGTATTCGT 6997
 Qy 403 LeuTyrrAlaThrAsnAspLeuThrLysTyrrAspSerGlyGlyGlnIleHisIleAlaPro 422
 Db 6998 GTGTGGGTCACTTATGAGTTAAAAAAGTCAATCGAAGGT-----ATGGGAAA 7045
 Qy 423 MetValLys-----GlyLeuGluGluThrAlaValAsnAspLeuIle 436

Db 7046 CCAGTGAATCGATATTCGAAATCGAGGGTTAGAGAAAACAGTTGTTTTCGAACCTGGTT 7105
 Qy 437 SerProSerGlyAlaProLeuLeuSerAlaLeuGlyAspLeuGlyPheThrHis 456
 Db 7106 CCCCCCATAGAGAAAGCCCTCTCTCTCGAATCGCCGATTCGGAGGATCCGTCAC 7165
 Qy 457 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThr 476
 Db 7166 GAAAGTTTGGACACCCACCACATCATCTATGTAC---AAACCCCTAAATGGAGCTCTCTT 7222
 Qy 477 SerValAspTyrAlaGluLeuAsnProSerIleIleValAlaGlySerPheAspPro 496
 Db 7223 GGTATAGCTTTTGGCTATCAGAAATTCAAAATTCGTGGCAGAGTGCATACATAC------ 7276
 Qy 497 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLeuAsnTrpPhe 516
 Db 7277 ---ACCTATCTCT-----TTCCTGTCTTACTCCGAGGATGGCGAATCAACTGGAGA 7324
 Qy 517 Gln---GlySerGluProGlyGlyValThrThrGlyGly-----ThrValAlaAlaSer 533
 Db 7325 GAGATAGAAACTGTCGGAGGGAATACTACGCGGAGGTTAAGTTTGGCAGTCTCG 7384
 Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 553
 Db 7385 AACGATGGAATAACACTGGTCTGGAGTCCAGCGAAT-----CATGAGGTAATTTGTTC 7438
 Qy 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla-----AsnAla 570
 Db 7439 AGCGATAAAGGAAAGCTGGAAAAGCCCAATAGTGTTCAGTTCCAGATTCCAACTAT 7498
 Qy 571 GlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeu-----SerAsnGly 588
 Db 7499 TTTCTCTGACTGATCTGTAAATCCGAGTAAATTTTACATCTTCGATTGGAAAAACGGT 7558
 Qy 589 ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 608
 Db 7559 GACTTCTCTAATAAGATGAGGAAAGAGTTTATGAAA---GGTGCTAAACTTCCT 7615
 Qy 609 SerSerGlyAlaValGlyValMetPheHisAla-----ValProGlyLysGlu 624
 Db 7616 TCTTTTGATAACTGGTGGGTATCGCTTTATTTCTTCCCTGTGTGGCTCTCTGACAGAGAA 7675
 Qy 625 GlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSerThrAsnGlyGlySer 643
 Db 7676 GGAGATATCTGTTAGCACTTCAATGGAATGGACTTTTACAGATCAAAAGATGGTGAATA 7735
 Qy 644 SerTrpSerAlaIleThrGlyValSerSerAlaValAlaValGlyPheGlyLysSerAla 663
 Db 7736 ACGTTTGAGGACTAGGAAAGCTTGATATAGCTACGTGATGGTTTCGGAGCCCAAAA 7795
 Qy 664 ProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla 683
 Db 7796 CCAGGTACTGACTATCCAGCAATTTTACCTGAACGGTATGGTGAATGGAGTTTATGGTATT 7855
 Qy 684 TyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspGlnHisGlnTyr 703
 Db 7856 TTCATGTCGACAGATAGGGGCAAAACATGGATGAGAAATTAACACAGATAGCATCAATTT 7915
 Qy 704 GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGly 723
 Db 7916 GGA---TGGATACATTACATGATAGAGATATGAATGAATTCGGAGAAATTTCTCTGGT 7972
 Qy 724 ThrAsnGlyArgGlyIleValTyrGlyAspIle 734
 Db 7973 ACGGAGGAGAGGAATCATCGTTGGTGAAGTG 8005

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 06:17:10 ; Search time 1138.43 Seconds
(without alignments)
3847.935 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATQPTWSNVAIGGGFVD.....YIGTNGRCIVGDIGGAPSG 740

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODE=frame+p2n.model -DB=spool/US09917376/runat_04102005.164327.26889/app_query.fasta_1.2524
-Q/cn2_1/USPTO.spool -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-DB=N Geneseq 16Dec04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPT=0 -UNITS=bits -START=1 -END=1 -MAIRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CN 1.1 1867 @runat_04102005.164327.26889 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04: *
1: geneseqn1980s: *
2: geneseqn1990s: *
3: geneseqn2000s: *
4: geneseqn2001as: *
5: geneseqn2001bs: *
6: geneseqn2002as: *
7: geneseqn2002bs: *
8: geneseqn2003as: *
9: geneseqn2003bs: *
10: geneseqn2003cs: *
11: geneseqn2003ds: *
12: geneseqn2004as: *
13: geneseqn2004bs: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4036	100.0	2869	10 ADD22922	Add22922 Acidother
2	4036	100.0	2869	10 AB277632	Ab277632 Nucleotid
3	4036	100.0	2869	12 AD052314	Ad052314 Acidother
4	2063	51.1	2823	8 ABV76941	Abv76941 Nucleotid
5	1579	39.1	2517	12 ADH51583	Adh51583 T reesei

6	1579	39.1	2710	12 ADH51580	Adh51580 Trichoder
7	1506	37.3	2849	10 ADD42060	Add42060 Trichoder
8	1139	28.2	3668	10 ADD24893	Add24893 DNA encod
9	1117	27.7	2367	12 ADH19119	Adh19119 Geotrichu
10	1117	27.7	2646	12 ADH19117	Adh19117 Geotrichu
11	1114	27.6	2481	12 ADH19123	Adh19123 Geotrichu
12	979	24.3	2268	13 ADR90297	Adr90297 Geotrichu
13	979	24.3	2274	13 ADR90301	Adr90301 Geotrichu
14	979	24.3	2755	13 ADR90295	Adr90295 Geotrichu
15	878.5	21.8	5698	10 ADD42054	Add42054 Trichoder
16	392.5	9.7	1103	3 AAF14988	Aaf14988 Trichoder
17	238	5.9	7407	8 ACA42281	Acc42281 Prokaryot
18	238	5.9	7407	9 ACC59398	Acc59398 Microbial
c 19	227.5	5.6	3129	11 ABD15269	Abd15269 Pseudomon
c 20	227.5	5.6	8211	11 ABD15052	Abd15052 Pseudomon
c 21	203	5.0	4188	11 ABD15170	Abd15170 Pseudomon
c 22	201.5	5.0	4131	11 ABD15169	Abd15169 Pseudomon
c 23	199.5	4.9	7305	8 ACA23399	Ac23399 Prokaryot
c 24	195	4.8	5295	8 ACA24001	Ac24001 Prokaryot
c 25	192	4.8	9903	8 ACA40315	Ac40315 Prokaryot
c 26	192	4.8	110000	4 AAI99682_04	Continuation (5 of
c 27	188.5	4.7	4399	13 ADS61968	Ads61968 Bacterial
c 28	184	4.6	3324	8 ACA38418	Ac38418 Prokaryot
c 29	184	4.6	110000	4 AAI99683_04	Continuation (5 of
c 30	182	4.5	11148	8 ACF39394	Acf39394 Mycobacte
c 31	182	4.5	69350	12 ADM98959	Adm98959 Diterpene
c 32	182	4.5	110000	4 AAI99682_37	Continuation (38 o
c 33	182	4.5	110000	4 AAI99683_37	Continuation (38 o
c 34	181.5	4.5	11679	9 ADA30090	Ada30090 DNA encod
c 35	180	4.5	2355	13 ADS62118	Ads62118 Bacterial
c 36	179.5	4.4	3222	8 ACA26562	Ac26562 Prokaryot
c 37	177.5	4.4	110000	4 AAI99682_30	Continuation (31 o
c 38	177.5	4.4	110000	4 AAI99683_30	Continuation (31 o
c 39	176.5	4.4	9855	10 ACF04821	Acf04821 M.licheni
c 40	176.5	4.4	51855	10 ACF04818	Acf04818 Melithiaz
c 41	176	4.4	32367	2 AAV35620	Aav35620 Human SHO
c 42	174.5	4.3	67459	8 ABQ77491	Abq77491 S. aurant
c 43	174	4.3	2855	4 ABL29005	Ab129005 Drosophil
c 44	174	4.3	5221	4 ABL19628	Ab119628 Drosophil
c 45	174	4.3	5302	4 ABL29004	Ab129004 Drosophil

ALIGNMENTS

RESULT 1
ADD22922
ID ADD22922 standard; DNA; 2869 BP.
XX
AC ADD22922;
XX
DT 15-JAN-2004 (first entry)
XX
DE Acidothermus cellulolyticus avicelase Aviii DNA.
XX
KW Aviii; cellulose reduction; agricultural biomass; municipal solid waste;
glycoside hydrolase; avicelase; ds; gene.
XX
OS Acidothermus cellulolyticus.
XX
PH Key Location/Qualifiers
FT CDS 1..2868
FT /*tag= a
FT /product= "Aviii"
XX
PN US2003108988-A1.
XX
PD 12-JUN-2003.
XX
PF 18-OCT-2002; 2002US-00155400.
XX
PR 28-JUL-2001; 2001US-00917376.
XX
PA (DING/) DING S.

PA (ADNE//) ADNEY W S.
PA (VINZ//) VINZANT T B.
PA (HIMM//) HIMMEL M E.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
XX WPI; 2003-810853/76.
DR P-PSDB; ADD22921.
XX
PT New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding AviiiI and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX
XX Claim 17; SEQ ID NO 2; 29pp; English.
XX
CC The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable AviiiI polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding AviiiI. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents DNA encoding the Acidothermus cellulolyticus
CC avicelase AviiiI.
XX
SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 1,17e-251 Length: 2869
Score: 4036.00 Matches: 740
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-917-376-3 (1-740) x ADD22922 (1-2869)

QY 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20
DB 139 GCGAGACTACGCGGTACACCTGGAGCAACGTCGGCGATCGGGGGCGCGCTTCGAC 198

QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 40
DB 199 GGGATCGTCTTCAATGAAGTGCCAGCGAATCTGTACGTGCGGACGACATCGGGGG 258

QY 41 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60
DB 259 ATGTATCGATCGGATGCCCAACCGCGCGGTGATCCCTTCTGGAATGGGTGGATGG 318

QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
DB 319 AACAAATGGGGGTACACGGCGTCTGCAGCATTCGGCGAGACCCCGCATCAATAACAA 378

QY 81 ValTrpAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
DB 379 GTATGGGCGCGTCCGATGTACACCAACAGCTGGGACCCAAACGACGAGCGATCTC 438

QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120
DB 439 CGCTCGTCTGATCAGGGGCGAACGTGGCNAATAAACGCCCTCGCTTCAAGCTTGGCGG 498

QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
DB 499 AACATGCCCGCGGTGGAATGGCGGAGCGCTTCGGTGGATCCAAACAAATGACACAT 558

QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
DB 559 CTGTATTTCCGGCCCGGAGCGGAAAGGGCTCTGGAGAAGCACAGATTCGGCGCGACC 618

QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
DB 619 TGGTCCCAGATGACCAACTTTTCCGAGCGTAGGACACGTATTCGAAATCCCACTGACAG 678

QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200
DB 679 ACCGGCTATCAGAGCGATATTCAAGGCGTCGTCGGTCTTCGACAAGTCTTCGTCA 738

QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
DB 739 TCGCTCGGGCAAGCGAGTAAGACCATTTTGTGGCGTGGCGGATCCCAATAATCCGGTC 798

QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyValaProThrGlyPhe 240
DB 799 TTCTGGAGCAGAGCGCGCGCGAGCTGCGAGCGGTGCGGGTGGCGGACCGCGCTTC 858

QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
DB 859 ATCCCGCACAGGGGTCTTTGACCCGGTCAACCGCTCTATATATGCGCACCAAGCAAT 918

QY 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280
DB 919 ACGGGTGGTCCGTATGACGGGAGCTCCGGCGAGCTCTGGAATTTCTCGGTGACCTCCGG 978

QY 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
DB 979 ACATGGACGCGAATCAGCCCGTACCTTCGACGGACACCGGCCCAACGACTACTTTGGTAC 1038

QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
DB 1039 ACGCGCTCACTATCGACCGCCGACACCCGACACGATATGTTGGTGGCAACCCAGATATCG 1098

QY 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340
DB 1099 TGGTGGCGGACACCATATCTTTTGGAGCACCGACCGCGGTGCGAGCTGGACGCGGATC 1158

QY 341 TrpAspTrpThrSerTyrProAsnArgSerLeuLeuArgTyrValLeuAspIleSerAlaGlu 360
DB 1159 TGGGATTTGGACGAGTTATCCCAATCGAAGCTTGTGCTTGTGTCATTTTCGGCGGAG 1218

QY 361 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380
DB 1219 CTTTGGCTGACCTTTCGGGTACAGCCGAATTCCTCCGTATCCAGTCCGAGCTCGGCTGG 1278

QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
DB 1279 ATGGATGAAGCGATGCAATCGATCGTTCAACTCTGATCGGATGCTCTACGGAACAGGC 1338

QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
DB 1339 GCGACGTTGTACGCAACAAATGATCTCAGAAAGTGGGACTCCGGCGGCCAGATTCAATC 1398

QY 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 440
DB 1399 GGGCGATGGTCAAGGATTTGGAGGAGACGCGGTAAACGATCTCATACCCCGCGCTCT 1458

QY 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThr 460
DB 1459 GGGCGCCCGCTCATCGCGCTCTCGGAGACTTCGGGGCTTCACCCACGCGGAGCTTACT 1518

QY 461 AlaValProSerThrIlePheThrSerProValPheThrThrThrGlyThrSerValAspTyr 480
DB 1519 GCGGTGCCATCGACGATCTTCAGTCCGGTGTTCAGCACCGCGCCACCGCTCGACTAT 1578

QY 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500
DB 1579 GGGGAATTGAATCCGTCGATCATCTGTCGCTGGAAGTTTCGATCCATCGAGCAACCG 1638

QY 501 AsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGlu 520
DB 1639 AACGACGACGCTCGGTTCTTCGACGAGCGCGGCAAGNACTGGTTCCAGGCGAGCGAA 1698

QY 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540
DB 1699 CCTGGCGGGGTGACGACGCGGCGCACCGTCCGCGCATCGCGCGAGCTCTCGTTCTGTC 1758


```
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSerSer 200
Db 679 ACCGGCTATCAGACGATATTCAGGGCTGCTCGGGTCCGCTTCGACAGTCTTCGTCA 738
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
Db 739 TCGCTCGGGCAAGCGAGTAAGACCATTTTGTGGCGTGGCGGATCCCAATAATCCGGTC 798
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240
Db 799 TTCTGGAGCAGACGCGCGCGACGCTGGCGCGGTGGCGCGACCGGCTTC 858
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
Db 859 ATCCCGCAGACAGGCGCTTTTGACCCCGTCAACACGCTCTATATATGACACAGCAAT 918
Qy 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280
Db 919 ACGGTCGTCGGTATGACGGGAGCTCCGGCGACGCTCTGGAAATTTCTCGGTGACCTCCGG 978
Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
Db 979 ACATGACGGCAATCAGCCGGTACCTTCGACGACACGGCCACGACTACTTTTGGTTAC 1038
Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
Db 1039 AGCGGCTCACTATCGACCGCGACACCGCAACAGATATATGTTGGCAACCCAGATATCG 1098
Qy 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgile 340
Db 1099 TGGTGGCGGACACCAATAATCTTTTCGGAGCACCGACCGCGGTGGACGTGGCGGATC 1158
Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
Db 1159 TGGGATTGGACGAGTATCCCAATCGAAGCTTGGGATATGCTTGACATTTTCGGCGGAG 1218
Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380
Db 1219 CCTTGGCTGACCTTCGGCGGTACACCGCAATCCTCCCGTACCGCAGTCCGAAGCTCGGCTGG 1278
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
Db 1279 ATGGATGACGATGGCAATCGATCCGTTCAACTCTGATCGGATGCTCTACGGAAACAGGC 1338
Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
Db 1339 GCGACGTTGTACGCAACAAATGATCTACGAAAGTGGGACTCCGGCGGCGCAGATTCAATC 1398
Qy 421 AlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProSer 440
Db 1399 GCGCCGATGGTCAAGGATGGAGAGACGGCGGTAAACGATCTCATCAGCCCGCGCT 1458
Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460
Db 1459 GCGCCCGCTCATCAGCGCTCTGGAGACCTCGCGGCTTCACCCAGCCGCGAGTTACT 1518
Qy 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
Db 1519 GCGGTGCCATCGACGATCTTCACGCTCACCGGTGTTCCAGCCCGCACCGCGTCCGACTAT 1578
Qy 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500
Db 1579 GCGGAATTTGAATTCGTCGATCATCTGCTTCGCGCTGGAAATTTTCGATCCATCCAGCAACCG 1638
Qy 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu 520
Db 1639 AACGACAGGACGCTCGCGCTTCGACACAGCGCGCAAGAACTGTTCCAAAGGACGCGAA 1698
Qy 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540
Db 1699 CCTGGCGGGTGCAGCAGCGGGCGGACCGCTCGCGCATCGCGCGGCTCTCGTTGCTC 1758
```

```
Qy 541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560
Db 1759 TGGGCTCCGCGCATCCGGTCAGCTGTGTGTGTACGCGATCGGATTTGGCAACTCTCTGG 1818
Qy 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580
Db 1819 GCTGCTTCGCAAGGTGTTCCGCCAATCCAGATCCGCTCAGACCGGTGTGAATCAAAG 1878
Qy 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600
Db 1879 ACTTCTATGCCCTATCCAATGGAACTTCTATCGAAGCAGCAGCGCGGTGACATTC 1938
Qy 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 620
Db 1939 CAACCGGTTCGCGCGCGGTCTTCCGAGCAGCGGTGCGGTGCTCATGTTCACGCGGTG 1998
Qy 621 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 640
Db 1999 CCTGAAAAGAAAGCGGATCTGTGGCTCGCTGCATCGAGCGGCTTTACCACTCAACCAAT 2058
Qy 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660
Db 2059 GCGCGCAGAGTGTGTCTGCAATCACCGCGGTATCTCCGCGTGAACTGGGATTTGTT 2118
Qy 661 LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyVal 680
Db 2119 AAGTCTGCGCGCGGTGCTCATACCAAGCCGCTTTTGTGTCGCGCACGATCGAGGCGTT 2178
Qy 681 ThrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGln 700
Db 2179 ACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTACTCATCAATGATGCCAG 2238
Qy 701 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720
Db 2239 CACCAATACGGAATTTGGGACAAGCAATCACCGGTGACCAAGCAATTTAGCGGGTG 2298
Qy 721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740
Db 2299 TACATAGCACGACGCGCGTGAATTTGTATACGGGACATTTGTGTGTCGCGCTCCGGA 2358
RESULT 3
AD052314
ID AD052314 standard; DNA; 2869 BP.
XX AC AD052314;
XX DT 15-JUL-2004 (first entry)
XX DE Acidothermus cellulolyticus avicelase III (AviIII) DNA.
XX KW Thermostable cellulase; avicelase III; AviIII; cellulose reduction;
XX KW biomass degradation; ethanol formation; industrial chemical;
XX KW fabric treatment; gene; ds.
XX OS Acidothermus cellulolyticus.
XX FH Location/Qualifiers
FT CDS 1..2868
FT /*tag= b
FT /product= "Acidothermus cellulolyticus avicelase III"
FT /note= "CDS contains translational exceptions"
FT /partial
FT /note= "No stop codon"
FT sig_peptide 1..306
FT /*tag= a
FT mat_peptide 307..2865
FT /*tag= c
FT /product= "Acidothermus cellulolyticus mature avicelase
FT III"
FT misc_feature 2869
FT /*tag= d
FT /note= "A string of unknown nucleic acid units"
XX
```


PN US2004038334-A1.
 XX 26-FEB-2004.
 XX 28-JUL-2001; 2001US-00917376.
 XX 28-JUL-2001; 2001US-00917376.
 XX (DING/) DING S.
 PA (ADNEY/) ADNEY W S.
 PA (VINZ/) VINZANT T B.
 PA (HIMM/) HIMMEL M E.
 XX
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;
 XX
 DR WPI; 2004-203224/19.
 DR P-PSDB; ADO52313.
 XX
 PT Novel thermostable AviiiI polypeptide of glycoside hydrolase family and
 PT isolated from Acidothermus cellulolyticus, useful for degrading treated
 PT biomass into simpler forms of carbohydrate.
 XX
 PS Claim 17; SEQ ID NO 2; 19pp; English.
 XX
 CC The invention relates to a thermostable cellulase enzyme, avicelase III
 CC (AviiiI) and its nucleic acid sequence. AviiiI is useful for reducing
 CC cellulose in a starting material. A thermostable AviiiI peptide is useful
 CC for degrading treated biomass into simpler forms of carbohydrate, which
 CC is used in the formation of ethanol or other industrial chemicals. It is
 CC also useful for treating fabrics to remove cellulose-containing stains.
 CC The present sequence is Acidothermus cellulolyticus AviiiI DNA.
 XX
 SQ Sequence 2869 BP; 546 A; 857 C; 899 G; 565 T; 0 U; 2 Other;
 Alignment Scores:
 Pred. No.: 1,17e-251 Length: 2869
 Score: 4036.00 Matches: 740
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0
 US-09-917-376-3 (1-740) x ADO52314 (1-2869)
 QY 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20
 DB 139 GCGAGCACTCACCGGTACACTGGAGCAACGTGGCGATCGGGCGCGGCTTTGTGCAC 198
 QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40
 DB 199 GCGATCGTCTTCAATGAAGTGCACCGGGAATCTGTAGTGGACGACATCGGGGG 258
 QY 41 MetTyrArgTyrAspAlaIleAsnGlyArgTyrIleProLeuLeuAspTrpValGlyTyr 60
 DB 259 ATGTATCGATGGATGGCCCAACGGCGGTGATCCCTCTTCGTGATTTGGGTGGATGG 318
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
 DB 319 AACAAATGGGGGTACAAACGGCGTGTGACATTTGGCGAGACCCGATCACTAACAAG 378
 QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
 DB 379 GTATGGCGCGCGTGGATGTACCAACAGCTGGGACCCCAACACGACGAGCGATTCTC 438
 QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly 120
 DB 439 CGCTGCTCTGTATCAGGGCGCAACGTGGCAATAAACGCCCTTCGCGTTCAAGTTGGCGGC 498
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAspAsnIle 140
 DB 499 AACATGCCCGGGCGTGGAAATGGGCGAGCGGCTTGGGTGGATCCAAACATGACACATT 558
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160

DB 559 CTGTATTTCGGCGCCCGAGCGCAAAAGGGCTCTGGAGAAGCACAGATTCCGCGCGCAC 618
 QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
 DB 619 TGGTCCCAGATGACGAACCTTTCCGAGACGTAGGACGATCATTTGCAAAATCCACTGACACG 678
 QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200
 DB 679 ACCGCTATCAGAGCGATATTCAAGGCGTCTGCTGGGTGCTTTGACAAAGTCTTCGTCA 738
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
 DB 739 TCGCTCGGGCAAGCGAGTAAAGACCATTTTGTGGCGGTGGCGGATCCCAATAATCCGCTC 798
 QY 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyValAlaProThrGlyPhe 240
 DB 799 TTCTGGAGCAGAGACGCGCGCGAGCTGGAGCGGTGCCGGGTGCGCGCCGCTTC 858
 QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
 DB 859 ATCCCGCACAGGGCGTCTTTGACCGGTCAACGACGTGCTCTATATTGCCACCAAT 918
 QY 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280
 DB 919 ACGGCTGGTCCGTATGACGGGAGCTCCGCGACGCTCGAAATCTCGGTGACCTCCGGG 978
 QY 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
 DB 979 ACATGGACGCGAATCAGCCGCTACCTTCGACGACGACGCGCAACGACTACTTTGGTTAC 1038
 QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
 DB 1039 AGCGGCTCACTATCGACCGCCAGCACCCCAACACGATTAATGGTGCACACCCAGATATCG 1098
 QY 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyValAlaThrTrpThrArgile 340
 DB 1099 TGGTGGCGGACACCATATATCTTCGAGCACCGAGCGGTGCGAGTGGAGCGCGATC 1158
 QY 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
 DB 1159 TGGGATGGACGAGTTATCCCAATCGAAGCTTGGATATGTGCTTGACATTTTCGGGGAG 1218
 QY 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTyr 380
 DB 1219 CCTTGGCTGACCTTCGGCGTACAGCGGAATCTCCGTACCCAGTCCGAGCTCGGCTGG 1278
 QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
 DB 1279 ATGGATGAACGATGGCAATCGATCCGTTCACTCTGATCGGATGCTCTACGGAACAGGC 1338
 QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
 DB 1339 GCGAGCTGTACGCAACAAATGATCTCACAAAGTGGGACTCCGCGCGCCAGATTCAATATC 1398
 QY 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSer 440
 DB 1399 GCGCCGATGGTCAAAGATTGGAGGAGCGCGGTAAACGATCTCATACGCCCGCGCTCT 1458
 QY 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThr 460
 DB 1459 GCGCCCGCTCATACGCGCTCTCGGAGACCTCGCGGGTTCACCCACGCGGCTTACT 1518
 QY 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
 DB 1519 GCGGTCCATCGACGATCTTCAGTCAACGGTGTTCAGACCGGACCGACGCTGACTAT 1578
 QY 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500
 DB 1579 GCGGAATTGAATCCGTCGATCATCGTTCGCGCTGGAAGTTTCGATCCATCCAGCCAAACCG 1638
 QY 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu 520

Db 1639 AACGACGACGCGCTCGCGTCTCTGCACAGACGCGCGCAAGAACTGGTTCCAGGCGAGCAA 1698
 Qy 521 ProGlyGlyValThrThrGlyGlyValThrValAlaAlaSerAlaAspGlySerArgPheVal 540
 Db 1699 CCTGGCGGGTGAAGACGCGGGCGCACCGTCCCGCATCGCGCGCTCTCGTTTCGTG 1758
 Qy 541 TrpAlaProGlyAspProGlyGlnProValValTrValAlaValGlyPheGlyAsnSerTrp 560
 Db 1759 TGGCTCCCGCGCATCCCGGTGAGCTGTGTGTAGCGAGTCGGATTGGCAACTCCTGG 1818
 Qy 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580
 Db 1819 GCTGCTTCGCAAGGTGTTCCGCCAATGCCAGATCCGCTCAGACCGCGGTGAATCCAAAG 1878
 Qy 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600
 Db 1879 ACTTTCATGCCCCATCAATGGAACCTTCATGAAGCACGACGCGCGGTGACATTC 1938
 Qy 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValAlaValMetPheHisAlaVal 620
 Db 1939 CAACCGGTGCGGCGCGGTCTCCGAGCAGCGGTGCGTGTGTATGTTCCACGCGGTG 1998
 Qy 621 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 640
 Db 1999 CCTGGAAAAGAGCGATCTGTGGCTCGCTGCATCGAGCGGCTTTACCACTCAACCAAT 2058
 Qy 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660
 Db 2059 GCGCGCAGCAGTTCGTCTGCAATCAGCGCGGTATCTCCGCGGTGAACGTGGGATTGCT 2118
 Qy 661 LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyVal 680
 Db 2119 AAGTCTGCGCGCGCGGTGCTATACCCAGCGTCTTTGTCTGTCGGCAGATCGAGCGGT 2178
 Qy 681 ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGln 700
 Db 2179 ACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGTACTGATCAATGATGACCCAG 2238
 Qy 701 HisGlnTyrGlyAsnTrpGlyAlaIleThrGlyAspHisAlaAsnLeuArgVal 720
 Db 2239 CACCAATACGGAAATGGGGACAAGCAATACCGGTGACCAACGGAATTTACGCGGGGTG 2298
 Qy 721 TyrIleGlyThrAsnGlyArgGlyTleValTyrGlyAspIleGlyGlyValAlaProSerGly 740
 Db 2299 TACATAGCACGAACGCGCGGTGGAAATGTATACGGGGACATTTGTGTGCGCCGCTCCGA 2358

RESULT 4

ABV76941
 ID ABV76941 standard; DNA; 2823 BP.
 XX AC ABV76941;
 XX DT 03-MAR-2003 (first entry)
 XX DE Nucleotide sequence of a xyloglucanase enzyme.
 XX KW Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;
 XX KW textile scouring; cellulose fiber processing; rattling; gene; ss.
 XX OS Jonesia sp.
 FH Key Location/Qualifiers
 CDS 1..2823
 FT /*tag= a
 FT /product= "xyloglucanase enzyme"
 FT sig_peptide 1..69
 FT /*tag= b
 FT mat_peptide 70..2820
 FT /*tag= c
 XX WO200277242-A2.
 XX 03-OCT-2002.
 XX PF 27-MAR-2002; 2002WO-DK000210.
 XX PR 27-MAR-2001; 2001DK-00000504.
 XX PA (NOVO) NOVOZYMES AS.
 XX PI Duffner F, Sjöholm C;
 XX DR WPI; 2003-092855/08.
 XX DR P-FSDB; ABB99489.
 XX PT New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,
 XX PT and endogenous to a bacterium, useful in the textile industry for
 XX PT improving properties of cellulosic fibers, yarn or fabric.
 XX PS Claim 11; Page 68-72; 76pp; English.
 XX CC The present sequence encodes a xyloglucanase enzyme, belonging to family
 CC 74 of glycosyl hydrolases. The enzyme is isolated from Jonesia sp.
 CC DSM14140. The enzyme is useful in processes for machine treatment of
 CC fabrics. It is also useful in the textile industry for improving the
 CC properties of cellulosic fibers, yarn, woven or non-woven fabric, and in
 CC a textile scouring process step. The xyloglucanase enzyme is also useful
 CC in the cellulose fiber processing industry for rattling of fibers such as
 CC hemp, jute, flax and linen. It is useful for preventing binding of
 CC certain soils to the xyloglucan left on the cellulosic material
 XX SQ Sequence 2823 BP; 665 A; 803 C; 764 G; 591 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5,01e-124 Length: 2823
 Score: 2063.00 Matches: 386
 Percent Similarity: 67.21% Conservative: 112
 Best Local Similarity: 52.09% Mismatches: 225
 Query Match: 51.11% Indels: 18
 DB: 8 Gaps: 9
 US-09-917-376-3 (1-740) x ABV76941 (1-2823)
 Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
 Db 130 TATAGCTGTCGAACTGTCAGATTGTGGCGCGGTTTGTGCTGTAATGTCCTTAAC 189
 Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45
 Db 190 CAGAAAGACCCAGGCTTGGTGTATGCCGACAGATATTTGGTGGAGCGTACCGGCTCAAT 249
 Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
 Db 250 GACTCCACCGGGCGTGGATCCCTTTACGGACCATATTTGGTTGGATGACTGGAGCCAC 309
 Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
 Db 310 TCCGGAATTCGTCTCTGGCCACTGACCTGTGTGACCAACCAACCGCGTGTACCTTGACGA 369
 Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
 Db 370 GGCACGTACTCCATGACTGGGATCCCAAAACGGTGTATTTTGGTTCGCGGACAAA 429
 Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
 Db 430 GCGGAGACTGGGAAAGACGATGTCGCTTCGTTGGGAGGCAACATGCCGGGTGCG 489
 Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
 Db 490 GGGATGGGAGAACGCTCGCGGTGACCCCAACAATAATAAAGTCTGTACTTTGGTGGC 549
 Qy 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
 Db 550 GAATCGGGTAATGGACTATGGAAGTCACTAGTACTACGGGAAAAACCTGGGGCAAGTGACG 609
 Qy 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly---TyrGln 184


```
Db 610 AGCTTCCCAATGCTGGTAACCTACGTGCGC-----GATGCTTCTGGTGGCGTACACC 660
Qy 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGln 204
Db 661 GCGCAAAACCAAGGTGTTGTGGGTAACTTTTGACCCCAATCAGCGAAAGCTGGTCAG 720
Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224
Db 721 ACAACACAAACGATCTACGTAGGTGTAGCAGACAAAGAGAACAGTTTATCGATCTAC 780
Qy 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLys 244
Db 781 GATGCGGGGCAACATGCAACGCTCTCGACACCCACCGGATTCCTGGCGCAAAAG 840
Qy 245 GlyValPheAspProValAsnHisValLeuTyrlleAlaThrSerAsnThrGlyGlyPro 264
Db 841 GGAGTGTCTGACCATAAAGGTCAGCAACTTACATCGCTACCTCAGATACTGTTGGCCG 900
Qy 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284
Db 901 TATGACGGGTGCAAGGGGATGTGGCGCCTAGATATTTCCAGTGGCCAGTGGACTCGA 960
Qy 285 IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr 304
Db 961 ATCAGCCCAATTCCTCTACCC-----TCATCGAATTCAGGCTTCGGGTATCTCGGTTAGCC 1017
Qy 305 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324
Db 1018 ATTGACCGCAAAACCCCTGACACCATCATGTTGTTTCCCAAGTCTCTGGTGGCCTGAC 1077
Qy 325 ThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThr 344
Db 1078 ATGTACGTATATCGCTCCACCGCGGGGAAACGTCGTGCGCTATCTGGAACTGAAC 1137
Qy 345 SerTyrProAsnAspSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThr 364
Db 1138 GGTTCACAGCCGCGTACCAAGCAGTACCAACCAAGTATCTCAGGGGCTCCCTGGCTTGAC 1197
Qy 365 PheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAla 384
Db 1198 TTGGAATACAGCAAGAGCTGAGCTAAGCTAAACCCCAACTCGGTGGATGACCCAGTCC 1257
Qy 385 MetalIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyr 404
Db 1258 TTGAGATTGACCCGCATAACTCCGACCGCTTCTTCTATGGCAGCGCGCGGCATCTAT 1317
Qy 405 AlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetVal 424
Db 1318 GCGGAACGAAACCTCACAACTGGGATAGGGCAAAAGTTGATATCACTGTCAAGGCC 1377
Qy 425 LysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 444
Db 1378 CARGGGATCGAAGAAACCGCAGCCCAAGACCTTGGCGCTCCCTCCCTGGAAACATGGCCTA 1437
Qy 445 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 464
Db 1438 TACTCAGCGCTGCGCAGACATCGGTGGCTTCCACCAAGATATCTCTCAGGTGCGCTAAC 1497
Qy 465 Thr-----IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483
Db 1498 AAGTACTACTACAAGAACCCACACCATGACACCGTGACCTCCATTTGACTTCGCCGAAAGC 1557
Qy 484 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503
Db 1558 AAACCGGCAACTGTGTGGTGCAGGCAAA-----TCATTCTGGGGAACCACTCTCA 1611
Qy 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523
Db 1612 TGGGTGGGTTCAACTGATGTGGCGAAACGTGGAAACCTGGCGCCACCGCATCAGGG 1671
Qy 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543
```

```
1672 GTTAAAGCCCGCGCTCCATCACTGTCAGCTAATGCTTCATCCATTGTCGGGCCCG 1731
Qy 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563
Db 1732 -----GAAGGTGCGCTCCTCGACCTTCCACCAACAGTGGCTCCTCGTGTCAACGGTG 1785
Qy 564 GlnGlyValProAlaAsnAlaGlnIleAerGserAspArgValAsnProLysThrPheTyr 583
Db 1786 AGCGGGCTTCTCACAATGCCAGGTGCGATCAGACCGAGTCAACGCTAATACTCTTTTAC 1845
Qy 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603
Db 1846 GCGTTCGTTGACGGAAGTTCTATCACTCCAAATGGTGGGCTTCCTTCACGCGCTCA 1905
Qy 604 Ala---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622
Db 1906 GCTTTCACGGGTTTCCACCCAGTGA-----AACGTTCTGTTTCCGTGCGTCCCGGCG 1959
Qy 623 LysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeuTyrHis 637
Db 1960 CGTCAAGTTCATCTCTGGCTTCTGCTGGAGTATCGGAAGCACCTATGGATGGCGG 2019
Qy 638 SerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 657
Db 2020 TCCACTGATGGCGGAAGAACTGGACCAAGGTATCTGGAGTTCAGGAAGCGCATGCCGTT 2079
Qy 658 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIle 677
Db 2080 GGTTCGGTAAAGCAACCACTAGTAGCGGATCCAGTGATTTTCACATCGCCCAAGATT 2139
Qy 678 GlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsn 697
Db 2140 GATGAGTTCGAGGAATTTTCGGTCTGACATGAGGCAAGCACTCGAAACGCATCAAC 2199
Qy 698 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717
Db 2200 GACAACCAACACAGTGGTATTGGACTGGCGCATCAATCAACCGGTGACCCAGATGTCGA 2259
Qy 718 ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAla 737
Db 2260 GGGCGGTCTACATTGTTACCAACGGTCGCGCATATTCTGTTGTTGATTCTCCTCACTCC 2319
Qy 738 Pro 738
Db 2320 CCG 2322
RESULT 5
ADH51583
ID ADH51583 standard; DNA; 2517 BP.
XX AC ADH51583;
XX DT
XX DT
XX DT
XX DE T reesei eg16 coding DNA sequence related to ethanol production.
XX EGVI protein; endoglucanase; ethanol production; biomass composition;
XX KW fermentation; cellulohydrolase; detergent production; softening agent;
XX KW cotton fabric; wood pulp degradation; sugar; gene; ds.
XX OS Hypocrea jecorina.
XX Key Location/Qualifiers
FH 1..2517
CDS /tag= a
FT /product= "Mature T reesei EGVI protein"
FT /note= "this sequence contains numerous translation
FT exceptions where the amino acid is given in the protein
FT sequence as J"
XX PN US2003113732-A1.
XX PD 19-JUN-2003.
```


XX 18-DEC-2001; 2001US-00026994.
 XX PF
 XX PA
 XX PR 18-DEC-2001; 2001US-00026994.
 XX PA (DUNN/) DUNN-COLEMAN N.
 XX PA (GOED/) GOEDEGEDEBUUR F.
 XX PA (WARD/) WARD M.
 XX PA (YAOJ/) YAO J.
 XX PI
 XX FI Dunn-Coleman N, Goedegebebur F, Ward M, Yao J;
 XX WPI; 2004-106460/11.
 XX DR P-PSDB; ADH51581.
 XX DR
 XX XX
 XX PT New substantially purified EGVI polypeptide with the biological activity
 XX PT of endoglucanase, useful for degrading biomass to ethanol.
 XX PS
 XX PS Example 1; SEQ ID NO 4; 26pp; English.
 XX CC This invention relates to a novel purified EGVI protein with the
 CC biological activity of an endoglucanase, comprising an amino acid
 CC sequence chosen from a fully defined sequence as given in specification.
 CC The protein is useful for producing ethanol, which involves contacting a
 CC biomass composition with an enzymatic composition containing the protein
 CC of the invention to yield a sugar solution, adding to the sugar solution
 CC a fermentative microorganism, and culturing the fermentative
 CC microorganism under conditions sufficient to produce ethanol, where the
 CC biomass composition may be optionally pretreated. The method further
 CC involves the addition of at least one endoglucanase or cellulohydrolase.
 CC The pretreatment is with a dilute acid. The protein of the invention is
 CC also useful for the production of a detergent composition which may be
 CC used as softening agent and for improving the feel of cotton fabrics and
 CC for degrading wood pulp into sugars. The present sequence is that of the
 CC coding DNA sequence which encodes the T reesei EGVI protein of the
 CC invention.
 XX SQ Sequence 2517 BP; 509 A; 791 C; 746 G; 471 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 9,01e-93 Length: 2517
 Score: 1579.00 Matches: 325
 Percent Similarity: 59.26% Conservative: 123
 Best Local Similarity: 42.99% Mismatches: 264
 Query Match: 39.12% Indels: 44
 DB: 12 Gaps: 20

US-09-917-376-3 (1-740) x ADH51583 (1-2517)
 QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24
 DB 61 TTTTCATGGAAGACGTCAAGCTCGCGCGCGCGCGCTTCGTCGCCGATCATCTTC 120
 QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44
 DB 121 CATCCAGACAAAGGGGTAGCATATGCACGACAGATATTGGCGGCTGTACCGCTC 180
 QY 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60
 DB 181 AAC---GCCGACGACTCATGACCGCCCTCAGGATGGATTGCTGATAATGTCGGCTGG 237
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
 DB 238 CACAACTGG-----GGCATCACGCTGTTCGCTTGTATCGCAGGACGATCAAAAG 288
 QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
 DB 289 GTGATGCCGACATCGGATGATATACGACAGCTGGATCCGAGTAATGGAGCCATCAIT 348
 QY 101 ArgSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120
 DB 349 CGCTCGTCAGACCGCGGCGCAACGTGGTCCCTTCACCAACTTGGCTTCCTCAAGTCGGGGT 408

QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
 DB 409 AACATGCCAGGACGCGGAGCGGAGAGCGTCTGGCTGTCATCCGCGCAACTCCACATC 468
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
 DB 469 ATCTACTTTGGTGTCTCGCTCAGGAACCGGCTCTCGAAAGTCTACGACGCGCGGTGACC 528
 QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
 DB 529 TTTTCCAAAGTCTCGTCTTCAGGCAACTGGGAGCTATCATCCAGACCCGAGTGATTC 588
 QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200
 DB 589 AACGGCTACAAACAGCAGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 648
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsnPro 219
 DB 649 ACGACCGGGGAGCCACGCTCTCGTATCTTTGTGGCAGCGCTGATAACATCATCTGCTCA 708
 QY 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239
 DB 709 GTCTATGTAGCAGCAAGATCCCGGCTCCAGTGGAGTGTCTGTACCGGGGAGCCAGGAAA 768
 QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259
 DB 769 TACTTTCCTCAAGCGCAAACTGCAGCCAGCAGCAGAGAGGCTTGTATCTGACCTATTC 828
 QY 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279
 DB 829 GATGCAACAGCGCGCTATGATGGCACACTTGGCTCAGTGTGGAGGTACGACATTCGAGG 888
 QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
 DB 889 GGAATTTGGAAGACATCATCCCTCTCTGTGATCAGATCTA-----TACTTTGGC 939
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
 DB 940 TTTGGCGGCTTGGCTCGATTTGCAAAAGCCAGGACCCCTGTGTGTTCTTCTTTGAAC 999
 QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339
 DB 1000 TCTTGGTGGCAGATGCTCAGCTGTTGGTTCGCTCGACCGACTCTGGGCAACATGGAGCCCG 1059
 QY 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359
 DB 1060 ATCTGGCGTGGCGGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAA 1119
 QY 360 GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSer----- 375
 DB 1120 GCACCGTGTATCAAGAACAACTTTATCGATGTGACGCGAGTCCCGTCCGATGCTCTC 1179
 QY 376 ---ProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394
 DB 1180 ATCAAGCGCTCGGCTGGATGATGATGCTCTCGAGATTGATCCCAACCCAGCAACCAAC 1239
 QY 395 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 414
 DB 1240 TGGCTCTACGGCAGCCGGAATGCAATCTTTGGCGGCGACGATCTCAACCACTGGGACAG 1299
 QY 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434
 DB 1300 CGCCACAAATGTCAATCAATCATCTGCAGACGGCATCGAGAAATTCCTCCGTCAGGAC 1359
 QY 435 LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454
 DB 1360 CTGGCTCTGCACCGCGGAGCGAGCTATTGGCGCGAGTCCGAGACCAACACGCGCTC 1419
 QY 455 ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal 471
 DB 1420 ACCTTTGCCAGCAGAAACGACCTCGGAGCATCCGCGCAGACGCTCTGGGCAACGCGCACA 1479
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491


```

Db 1480 TGGGCCACCTCGACGAGCGTGCATCGCCGGGAACCTCGGTCAAGAGCGTGGTCCGGCTC 1539
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
Db 1540 GGCACACCCCGCGCACCAACAG-----GTGGCCATCTCGTCCGACGGC 1584
Qy 512 GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyValAla 531
Db 1595 GCGCGGAGTGGAGCATCGACGCGGCCACAGCTCATGAACGGGGGACGGTGGCC 1644
Qy 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551
Db 1645 TATTCGCGCAGCGCACGACATCTCTGTCGACCGCTCGTCCGGC-----GTG 1695
Qy 552 TyrAlaValGlyPheGlyAsnSerTrpAlaSerGlnGlyValProAlaAsnAlaGln 571
Db 1696 CAGCGCTCGAGTTCCAGGGCAGCTTTGCTCTCGAGCTCGCGCGGGCGCGCTC 1755
Qy 572 IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591
Db 1756 ATCCGCTCGGACAAAGACCAACAGCGCTCTCTACGCGCGCTCCGATCGACCTTTTAC 1815
Qy 592 ArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611
Db 1816 GTCAGCAGGACACCGCGCAGCAGCTTC-----ACGCGCGG-----CCCAAGCTGGGC 1863
Qy 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629
Db 1864 AGCGCAGGAGCATCGGGATATCGTCTCACCGCACCGCGGGGACCTTGTATGTC 1923
Qy 630 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 648
Db 1924 TCGACCGACGTCGGCATATTCGCTCCACAGCTCGGGCAGACCTTTGGCCAAAGTCTCC 1983
Qy 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668
Db 1984 ACCGCTTGACCAACACCTACAGATCCCTCGGTGGTGGGTCTCA-----GGCTGACACG 2040
Qy 669 ProAlaValPheValValGlyThrGlyValThrGlyValAla-----TyrArgSer 686
Db 2041 ---AACCTGTATGCTTCGGCACC-----GGCCGCTCAGGGCTCGCTCTACGCCAGT 2091
Qy 687 AspAspCysGlyThrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrp 706
Db 2092 GGAGACAGCGCGCTCTCGACGACATCCAGGGCTCCAGGGCTTCGGCTCCATCGAC 2151
Qy 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGly 726
Db 2152 AGCACCAAGGTGCGCGGACGCGGACGCGCGGCAAGTCTAGTGGGACCAACAGGC 2211
Qy 727 ArgGlyIleValTyr-----GlyAspIleGlyAlaProSerGly 740
Db 2212 CGGGCGCTTTTACGCTCAGGGAACCTCGCGCGGCGGCGGCGG 2259

```

RESULT 6

```

ADH51580
ID ADH51580 standard; cDNA; 2710 BP.
XX
AC ADH51580;
XX
XX
DT 25-MAR-2004 (first entry)
DE Trichoderma reesei eg16 cDNA sequence related to ethanol production.
XX
KW EGVI protein; endoglucanase; ethanol production; biomass composition;
KW fermentation; cellobiohydrolase; detergent production; softening agent;
KW cotton fabric; wood pulp degradation; sugar; gene; ss.
XX
OS Hypocrea jecorina.
XX
FH Location/Qualifiers
CDS 98..2557

```

```

/*tag= a
/product= "Mature T reesei EGVI protein"
/Note= "This sequence contains numerous translation
exceptions where the amino acid is given in the protein
sequence as J"

US2003113732-A1.
19-JUN-2003.
18-DEC-2001; 2001US-00026994.
18-DEC-2001; 2001US-00026994.
(DUNN/) DUNN-COLEMAN N.
(GORD/) GOEDEGEBUUR F.
(WARD/) WARD M.
(YAOJ/) YAO J.
Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;
WPI; 2004-106460/11.
P-PSDB; ADH51581.
New substantially purified EGVI polypeptide with the biological activity
of endoglucanase, useful for degrading biomass to ethanol.
Example 1; SEQ ID NO 1; 26pp; English.
This invention relates to a novel purified EGVI protein with the
biological activity of an endoglucanase, comprising an amino acid
sequence chosen from a fully defined sequence as given in specification.
The protein is useful for producing ethanol, which involves contacting a
biomass composition with an enzymatic composition containing the protein
of the invention to yield a sugar solution, adding to the sugar solution
a fermentable microorganism, and culturing the fermentative
microorganism under conditions sufficient to produce ethanol, where the
biomass composition may be optionally pretreated. The method further
involves the addition of at least one endoglucanase or cellobiohydrolase.
The pretreatment is with a dilute acid. The protein of the invention is
also useful for the production of a detergent composition which may be
used as softening agent and for improving the feel of cotton fabrics and
for degrading wood pulp into sugars. The present sequence is that of the
cDNA sequence which encodes the T reesei EGVI protein of the invention.
XX
SQ Sequence 2710 BP; 571 A; 840 C; 782 G; 517 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9,81e-93 Length: 2710
Score: 1579.00 Matches: 325
Percent Similarity: 59.26% Conservative: 123
Best Local Similarity: 42.99% Mismatches: 264
Query Match: 39.12% Indels: 44
DB: 12 Gaps: 20

US-09-917-376-3 (1-740) x ADH51580 (1-2710)
Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAspGlyIleValPhe 24
Db 101 TTTTCATGGAAGAACGTCGAAGCTCGCGCGGGCGGCTCGTCCCGGCATCATCTTC 160
Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrp 44
Db 161 CATCCCAAGACAAAGGCGTAGCATATGCACGACAGATATTCGCGGCTGTACCGGCTC 220
Qy 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTyr 60
Db 221 AAC---GCCGAGACTCATGGACCGCGCTACCGGATGGGATTCGTGATATGCGGCTGG 277
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 278 CACAACTGG-----GGCATCGACGCTGTGTGGCTTCATCCGAGGACGACAAAG 328

```


KW Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;
 KW xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;
 KW ethanol production; detergent composition; fabric treatment;
 KW textile treatment; gene; db.

XX Hypocrea jecorina.

OS Location/Qualifiers

Key CDS
 59..2743
 /tag= b
 /product= "Family 74 xyloglucanase"
 /note= "Contains 3 introns"

exon
 59..381
 /tag= a
 /number= 1

intron
 382..440
 /tag= c
 /number= 1

exon
 441..1227
 /tag= d
 /number= 2

intron
 1228..1277
 /tag= e
 /number= 2

exon
 1278..2664
 /tag= f
 /number= 3

intron
 2665..2723
 /tag= g
 /number= 3

exon
 2724..2743
 /tag= h
 /number= 4

WO2003089598-A2.

30-OCT-2003.

17-APR-2003; 2003WO-US011831.

19-APR-2002; 2002US-0373987P.

(NOVO) NOVOZYMES BIOTECH INC.

Michael R, Zaretsky E, Haas J;

WPI: 2003-845528/78.

P-PSDB; ADD42055.

New polypeptides having Family 74 xyloglucanase activity, and encoding nucleic acid molecules, useful for degrading cellulose- and hemicellulose -containing biomass to ethanol or as a detergent.

Claim 1; Fig 1; 96pp; English.

The invention relates to a Family 74 xyloglucanase (ADD42055) from the fungus *Trichoderma reesei* (Hypocrea jecorina), and nucleic acids encoding it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic linkages in the backbone of xyloglucan to xyloglucan oligosaccharides. The invention also relates polypeptide sequences at least 70% identical to the enzyme, expression vectors and host cells comprising a nucleic acid of the invention, the recombinant production of the enzyme, and mutant enzymes and the nucleic acids encoding them. The xyloglucanase of the invention can be used in the degradation of cellulose- and hemicellulose-containing biomass to produce ethanol. It can also be used in a detergent composition for treating fabric during a machine washing cycle. The present sequence represents the gene encoding *Trichoderma reesei* Family 74 xyloglucanase. Note: A comparison of this sequence with that shown in the sequence listing (also referred to as SEQ ID NO:1) shows that the present sequence is half the length of the sequence listing SEQ ID NO:1 (ADD42054), as the sequence listing SEQ ID NO:1 contains alternate 60 bp blocks of sense strand and complementary strand sequence.

XX Sequence 2849 BP; 594 A; 882 C; 822 G; 551 T; 0 U; 0 Other;
 SQ Alignment Scores:
 Pred. No.: 5.49e-88 Length: 2849
 Score: 1506.00 Matches: 329
 Percent Similarity: 56.60% Conservative: 121
 Best Local Similarity: 41.38% Mismatches: 260
 Query Match: 37.31% Indels: 87
 DB: 10 Gaps: 23

US-09-917-376-3 (1-740) x ADD42060 (1-2849)

QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24
 Db 119 TTTTCATGGAAGAAGTCAAGCTCGCGCGCGGGCTTCGTCCCGGCATCATCTTC 178
 QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44
 Db 179 CATCCCAAGACAAAGGCTAGCATATGCAGAACAGATATTGGCGGGCTGTACCGCCTC 238
 QY 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60
 Db 239 AAC---GCCGACACTCATGTCGCGCGCGCTCACGGATGGGATTGCTGATTAATGCGCGCTGG 295
 QY 61 AsnAsnTrpGlyTyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnLys 80
 Db 296 CACAACCTGG-----GGCATCGACGCTGTTCCTGTATCCGACGAGCATCAAAAG 346
 QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp----- 93
 Db 347 GTGTATGCGCAGTGGCATGTATAGAACACAGCTGTC-TGTGATGTCTCTCAGATCTAG 405
 QY 94 -----ProAsnAspGlyAlaIleLeu 100
 Db 406 ACCTATGATTGGACGCTGACATTGGCCATATATAGGGATCCGAGTAATAGGAGCATCAT 465
 QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120
 Db 466 CGCTCGTCAGACGGCGCGCAACGTGGTCTTCACCAACTTGCCTTCAAGTCGGGGGT 525
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
 Db 526 AACATGCCAGAGCGCGGAGCGCGAGAGCGTCTGGCTGTGATCCGCGCAACTCCAAATC 585
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160
 Db 586 ATCTACTTTTGTGCTCGCTCAGGAAACGGCCTCTGGAAGTCTACGGACGGCGGTGACC 645
 QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
 Db 646 TTTTCCAGGTCTCGTCTTCAGGCAACTGGGACGCTACATCCAGACCCGAGGTATTC 705
 QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200
 Db 706 AACGGCTACAACAGCGCAGCAAGCAAGCACTATGTGGGTACGTTCGACTCAACCGAGCAG 765
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsnPro 219
 Db 766 ACGACCGGGGAGCCACGCTCGTATCTTTTGTGGCAGCGCTGATAACATCATCTGCTTCA 825
 QY 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239
 Db 826 GTCTATGTGAGCAGCAATGCGCGCTCCAGCTGAGTGTGTACCGGGCGAGCGGAGAA 885
 QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259
 Db 886 TACTTTCTTCACAAGCGCGAAACTGTCAGCCAGCAGAGAGGCCTTGTATCTGACCTATTC 945
 QY 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279
 Db 946 GATGGCACAGGGCCGTATGATGACCACTTGGCTCAGTGTGGAGGTACGACATTCGACGG 1005

CC antisense nucleic acid sequence. The nucleic acid sequence of the gene
 CC encoding the protein is a recombinant sequence having at least one
 CC mutation as compared to the wild-type gene encoding the protein. The
 CC transgenic cell or microorganism are useful for producing xanthan gum,
 CC which are useful for providing formulations and properties, such as long-
 CC term suspension and emulsion stability in alkaline, acid, and salt
 CC solutions, temperature resistance, and pseudoplasticity. The present
 CC sequence encodes an enzyme relating to the present invention.
 XX
 SQ Sequence 3668 BP; 570 A; 1181 C; 1292 G; 625 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,056-64 Length: 3668
 Score: 1139.00 Matches: 267
 Percent Similarity: 50.85% Conservatives: 120
 Best Local Similarity: 35.09% Mismatches: 295
 Query Match: 28.22% Indels: 79
 DB: 10 Gaps: 21

US-09-917-376-3 (1-740) x ADD24893 (1-3668)

QY 1 AlaThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20
 DB 1217 GCACGTCCGGCCCTACAGTGGCGAGTGTGCCATTGGCGGTTCGTACC 1276
 QY 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40
 DB 1277 GGTGTGCTGTTTCCATCCCGCGAAGCGTGTCTGGCTATGCGCAGCAGATGGGTGGC 1336
 QY 41 MetTyrArgTTPAspAlaAlaAsnGlyArgTTPleProLeuLeuAspTTPValGlyTTP 60
 DB 1337 GCGTACCGTGGGATGCGCAGCGCAGTGGAGTGGACCGCTGACCGACTGGTGGCGCT 1396
 QY 61 AsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
 DB 1397 GAGCATGGAACTGATGGGATCGACGATTCGCGTTCGACCCCGCATCGGATGG 1456
 QY 81 ValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyAlaIleLeu 100
 DB 1457 CTGTATCTGGCGCGCCACCTATATGTCATGAA---CGCGCGGCACTCGCGCAGTGTG 1513
 QY 101 ArgSerSerAspGlnGlyAlaThrTTPGlnIleThrProLeuProPheLysLeuGlyGly 120
 DB 1514 CGTCTGTTCAACCGCGCGCGCAGTTCGAGCGTGGCGACCTGCGGTTTAAGTGGGTGT 1573
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAenIle 140
 DB 1574 AACAGCTGGCGCGCGCCCAATGTCGAGCGCTGGCGGTGACCCGCGCATGGCGCGCAC 1633
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTTPArgSerThrAspSerGlyAlaThr 160
 DB 1634 CTGCTCTGGCTCGCGCGATGCC---GGCTGTGGCGTAGCGACGATCGCGCGCGCAC 1690
 QY 161 TTPSerGlnMetThrAsnPheProAspValGlyValTTPIleAlaAsnProThrAspThr 180
 DB 1691 TGGCGCAAGTGGCGTGGTTCGCGACCGCGCGCTGCGCGTCCACCGCGCGCAATCAT 1750
 QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTTPValAlaPheAspLysSerSer 200
 DB 1751 GTTGGCGCGGACGACGCGGTG---GGGATCGCTTTGTGTGTGTCAGCGCAGCATGGC 1807
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
 DB 1808 AACACGGCTCGCCACACCGCGCATCTAGTGGGCTGTCCACCGACAGACCGCTG 1867
 QY 221 PheTTPSerArgAspGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe 240
 DB 1868 TANTGTCCGAAGATGCCGCCCGCAGTGTGGCACCCTGGTGGCGGCAACCCACGCGCCTG 1927
 QY 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
 DB 1928 CGCCCGAGCCACATGCGCGCGCGCAGCATGGGCAC---TGGTATCTGAGCTATGGCGAC 1984

QY 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTTPIlePheSerValThrSerGly 280
 DB 1985 CAGCCCGCGCGGACCTGATGCCCGGGGAGCCTTGTGCAATTCACGCGGCGCACAGGG 2044
 QY 281 ThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
 DB 2045 CGCTGGCGTGGATCAGCCCGATTCGCGCAG---CCAGCCAGTGGCGCATGGAATTCGCTGG 2101
 QY 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
 DB 2102 GGTGGGTGGCGGTGGATCCGCAACATCCGAGGTGCTGCTGGCCAGCAGCTTCGCCCGT 2161
 QY 321 TTPTTPProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArgIle 340
 DB 2162 CGCAGCGCGCGGACGAGCTGATCGCAGCTGGATGGTGGCAAGCACTGGAGCGCGTGG 2221
 QY 341 TTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
 DB 2222 -----TTGGCGCATGCGGTGTTCGATCAGCGCGCGC 2254
 QY 361 ProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTP 380
 DB 2255 CCGTGG---ACGCGACATGCCAGCGCGCAC-----TGG 2284
 QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
 DB 2285 ATGGGG---GCCTGGCGATCGATCCGTCGACGGCAACCATCGCTGTTTCGTGACCGCG 2341
 QY 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIle 420
 DB 2342 TACGGCATCTGGGCTCGCGCAATCTG-----CAGGATTTCCGCC 2380
 QY 421 AlaPro-----MetValLysGlyLeuGluThrAlaVal 432
 DB 2381 GCACCGCAGCGCGCTGCGAGTGGTGTCCAGGACCGTGGGTGGAGAAACCGTGGCG 2440
 QY 433 AsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 452
 DB 2441 CTGGACCTCTCAGCCCGATGGCTGGCGCATCTGCTCAGCGCGCTCGCGGATATCGAC 2500
 QY 453 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 472
 DB 2501 GGCCTTCGGCATGACGACCTGGACCGCGTG---CAGTTGCAGTACGCGCGCCACCGGTG 2557
 QY 473 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 492
 DB 2558 ACCAATGGCGAAAGCATCGATGCGCGCGCAGCGCGCGCGTGGTGGTGGCGCGCGT 2617
 QY 493 SerPhe-AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyG 512
 DB 2618 ACCGTGCGCGCGCGCGCAACAGAAATCCGCGCGC---TGATTTAGCGGTGATGGCG 2674
 QY 512 YLysAsnTTPPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAl 531
 DB 2675 CAAGCAATGGATCGCTGCTTCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCATTCG 2734
 QY 531 aAlaSerAlaAspGlySerArgPheValTTPAlaProGlyAspProGlyGlnProValVa 551
 DB 2735 CATTTGGTGGCATGCGCGCGCGGTGTGTGGCGACCGCGCGCGCGCGCAAT-----TG 2788
 QY 551 lTyrAlaValGlyPheGlyAsnSerTTPAlaAlaSerGlnGlyValProAlaAsnAlaGl 571
 DB 2789 GCGCAGCTCCGATCTTCGGCGCGCAGTGGCAGCGGTGTGAACGCCCTCGCGCAACCCCGCTG 2848
 QY 571 nIleArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyTh 589
 DB 2849 GGTGGTGACCAATCGGTGGACGCGCGCTGATGCGGTGATGTCGGTGGATGTCGCGCGCGCA 2908
 QY 589 rPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu----- 607
 DB 2909 GCTGTACGAGAGCAGCAGCATGCGCGCGCGCAGTTTCCTG-----GCGACCGCTGTACAGGT 2962
 QY 608 -----ProSerSerGlyAlaVala 613

Db 2963 GGGCAGCCCGCGCGATGAGCGACCCCGCGCGAGCTGCGTCCGACCGCGTGGCGGC 3022
 Qy 613 lGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaSerSe 633
 Db 3023 AGGCGTGTGTATGTCGCGACCCCGGCAAG----- 3053
 Qy 633 rGlyLeuTrpHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSe 653
 Db 3054 -GGTGTGATGCGTGGCAGGAGGT-----GCGTGCAGTGTCTCTACAGCGCGGCGA 3106
 Qy 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTrpProAlaValPheVa 673
 Db 3107 AGCAGCTCGTGTGCGATCGCAAGCGCTGCGCGCGGTGCGCGCGTGTATCT 3166
 Qy 673 lValGlyThrIleGlyValThrGlyValArgSerAspAspCysGlyThrTr 693
 Db 3167 GCGCGCGCGTGCAGGCGTGGATGGGTGTTCGCTCCGACGATGCGCGCGCGAGTG 3226
 Qy 693 pValLeuIleAsnAspAspGlnHisGlnTrpGlyAsnTrpGlyGlnAlaIleThrGlyAs 713
 Db 3227 GCAGCGCATCAACGATGAGCGCGCCGCTTCGGGCGC---CCGTACAGCGTGACCGGTGA 3283
 Qy 713 pHisAlaAsnLeuArgArgValTrpIleGlyThrAsnGlyArgGlyIleValTrpGlyAs 733
 Db 3284 TCCGCGCATTCGCGCGGTGTGTCTTCGCCACCGCGCGCGGTATTTCTTACGCGGA 3343
 Qy 733 p 733
 Db 3344 T 3344
 RESULT 9
 ADH19119
 ID ADH19119 standard; DNA; 2367 Bp.
 AC ADH19119;
 DT 11-MAR-2004 (first entry)
 DE Geotrichum sp M128 xyloglucan oligosaccharide-degrading enzyme mat DNA.
 KW xyloglucan oligosaccharide degradation; genetic engineering; ds; gene;
 KW mature.
 OS Geotrichum sp. M128.
 FH Key Location/Qualifiers
 FT 1. .2367
 FT CDS /*tag= a
 FT /partial
 FT /product= "Geotrichum sp M128 xyloglucan oligosaccharide-
 FT degrading enzyme mature protein"
 FT /note= "No start or stop codon"
 XX
 PN EP1350844-A2.
 XX
 XX 08-OCT-2003.
 XX
 XX 25-MAR-2003; 2003EP-00251866.
 XX
 XX 25-MAR-2002; 2002JP-00083433.
 XX
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX
 XX Yaoi K, Mitsuishi Y;
 XX
 XX WPI: 2004-100948/11.
 XX P-PSDB; ADH19120.
 XX
 XX Novel xyloglucan oligosaccharide-degrading enzyme with a different
 XX degradation mechanism from known enzymes.
 XX
 XX Claim 8; SEQ ID NO 13; 39pp; English.

XX The invention relates to a novel xyloglucan oligosaccharide-degrading
 CC enzyme with a different degradation mechanism from known enzymes. The
 CC primary structure and polynucleotide structure of the xyloglucan
 CC oligosaccharide-degrading enzyme provided by the present invention allows
 CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation
 CC activity to be prepared at a low cost through a genetic engineering
 CC process. The current sequence is that of the Geotrichum sp. M128
 CC xyloglucan oligosaccharide-degrading enzyme mature DNA of the invention.
 XX
 SQ Sequence 2367 BP; 471 A; 797 C; 719 G; 380 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 6.5e-63 Length: 2367
 Score: 1117.00 Matches: 281
 Percent Similarity: 50.37% Conservative: 124
 Best Local Similarity: 34.95% Mismatches: 279
 Query Match: 27.68% Indels: 120
 DB: 12 Gaps: 31
 US-09-917-376-3 (1-740) x ADH19119 (1-2367)
 Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
 Db 10 TACGAGTTTCAAGAATGTCCGATCGCGCGCGGTATACATTACCGGATTCGCGCAC 69
 Qy 26 GluGlyAlaProGlyIleLeuTrpValArgThrAspIleGlyGlyMetTrpArgTrpAsp 45
 Db 70 CCAAGACCAAGACCTCTGTACCGCGCGACGACATTTGGCGCGGTACCGCTGGGAC 129
 Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTrp 65
 Db 130 GCAGCGACGTCCAAGTGGATCCGCTCAACGACTTTATCGAGCGCGACGACATGAACATT 189
 Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaVal 85
 Db 190 ATGGGCACCGAGTCGATCGCTGACGCCCAACACCCCGCAGCGCTGTACCTCGCGCAG 249
 Qy 86 GlyMetTrpThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104
 Db 250 GGGCGCTATGCGGCGACGAGTGG-----GGCGCGTTCCTATGTGTCGAGAC 297
 Qy 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124
 Db 298 CGCGCCAGTCGTTTACATCTACAGTCGCGCTCCGATCGCGCGCACACGACATGGGA 357
 Qy 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTrpPheGly 144
 Db 358 CGCAACAATGCGAGCGCTCGCTGTCAACCGCTTCAACTCGAAGAGGTCTGGATGGGT 417
 Qy 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
 Db 418 ACAGGTACA---GAGGGTATCTGGAAGAGTTGGACCGCGCCCAACACCTGGCAACACGTC 474
 Qy 165 ThrAsnPheProAspValGlyThrTrpIleAlaAsnProThrAspThrThrGlyTrpGln 184
 Db 475 ACGTCCATCCCGACGCGTTC-----ACCAACGGTATCGGATACACG 516
 Qy 185 SerAspIleGlnGlyValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204
 Db 517 TCG-----GTCATTTTCGACCCG-----GAA 537
 Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224
 Db 538 CGTAATGCGACCATCTACCGCGCGCGCTGCGCGCGAGGGC---ATGTACGTTCAGCGAC 594
 Qy 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241
 Db 595 GACGCGGTGTCTCGTGGGAGCCAGTGGCGCGCGCGTCCAGCTCGCTCAACAGGACC 554
 Qy 242 -----ProHisLysGlyVal 246
 Db 655 ACGGCGCGTTCGCCGACAGAGCCCGGTGATCGCGCGCGCGCATGAAGTCGCT 714

QY 247 PheaspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266
 DB 715 CTCACCCCC-----AACTTCCTCTACGTGACTTACGCCGACTACCTCGGTCCATGGGCG 768
 QY 267 GlySerSerGlyAspValTyrPheSerValThrSerGlyThrThrArgile--- 285
 DB 769 GTCAGTTCGGCGAAGTTCGGCGCAGNACCGCACCTCGGGCGCTGGGACGACATTACT 828
 QY 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
 DB 829 CCCCGCTCGCAACTCGTCGCTGCCCGCTACAAACACACGAGCTTCCTCGCGCGGA 888
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
 DB 889 TTTTCGGTCTCAGCGTCGACGCGACCAACCCCAACCGCTCGTCGTCTATCACC---CTC 945
 QY 320 SerTyrTyrProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336
 DB 946 GACCGGACCCCGGACCGCGCTCGACAGCATCTACCTCTCACCAGATCGCGCGCGACC 1005
 QY 337 TrrThrArgIleTrrAspTrrThrSer-----TyrPro 347
 DB 1006 TGGAGGAGCTCACCACCTCTCGTCCCGTCCAACTCGAAGTAACCTGGGCGCACCG 1065
 QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrrPheThrPheGlyVal 367
 DB 1066 ACTAACGCGCGCGGTAC---AAGACGCGCACGCTGTTCGTGCTCGACTTCAACAC 1122
 QY 368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379
 DB 1123 GGTCCCAAGTGGGGGAGATACGGTGGCGGCACGCTCTCCGTGTCGAGAGGACTGG----- 1293
 QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399
 DB 1183 TGTGGATGAGCGCTGTCTTATCGATCGTTCACCCCGACGACTGTATGATGACGACG 1242
 QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 415
 DB 1243 GGGCGGACCATCTGGCGGACCGACGCTCTCCGTGTCGAGAGGACTGG----- 1293
 QY 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 431
 DB 1294 -----GCGCGGAGCTGGTACCTCCAGATCGACGCTATCGAGAGAAATGCG 1338
 QY 432 ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451
 DB 1339 ATCTGTGCTCGCTCGCTCCGAGAGCGCGCGCTCTCTGTGGGATCGGTGACATT 1398
 QY 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471
 DB 1399 AGCGGATGAGACGACGACCTCACAAG---CCCCAGAAGATGTTTGGTGGCGCCAG 1455
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491
 DB 1456 TTTCTCAACCTCGACAGCATCGACGCTGGGGCAACTTCCCAAGCTGTGCTCGCGCC 1515
 QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
 DB 1516 GATATCTCGGGACACGAGTACGACAGCGGTGGCGCGGTGGTGGTACCGGACTGACGCG 1575
 QY 512 GlyLysAsnTrrP-----PheGlnGlySerGluProGlyGlyValThrThr----- 526
 DB 1576 GGNAGCGGTGAGCATCTTCCCTACCTGCTCCCTGATGAAACGCGACCTACACAG 1635
 QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrrPalaPro-----Gly 544
 DB 1636 GGCAGCAGCATGTCAGTCAGCGGACGCGGACGACAGATGCTGTGTGTCGACCAAGCTTGAC 1695
 QY 545 AspProGlyGlnProValTrrValAlaValGlyPheGlyAsnSerTrrPalaAlaSerGln 564
 DB 1696 GAGCAGGCTCGGGACCGGTGGTACTGCGACGACTATGGCAAGACGTGGTCT----- 1746

QY 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578
 DB 1747 ---GTTCCCGTGGCGACTGGAAGCCCGAGACTGCAATGTGTCTTCGACAAAGTCCAG 1803
 QY 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 598
 DB 1804 GATGCACGTTCTACGCTACCGATCGCGCAAGTTCTTCGTCTCGACCGGACGCGGGAAG 1863
 QY 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
 DB 1864 TCGTATCGCCCAAGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 1917
 QY 619 AlaValProGlyLysGluGlyAspLeuTrrPLeu---AlaAlaSerSerGlyLeuTrrHis 637
 DB 1918 GTGAACCCCTGGGTGGCGGCGACGCTGTGGGTGCTGTTCCTCCGAGGGCGGTCTCTCCAC 1977
 QY 638 SerThrAsnGlyGlySerSerTrrPLeuAlaIle---ThrGlyValSerSerAlaValAsn 656
 DB 1978 TCGACCGACTTGGCGCTCTGTTTCAGAGGTAGGTACCGCAACGCGACCTCGTGAGC 2037
 QY 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTrrProAlaValPhe 672
 DB 2038 GTCCGCGCCCAAGTCCAAAGTCGACGCGCAAGAGGTAGCGCGCTCCGCGTCTTC 2097
 QY 673 ValValGlyThr-----IleGlyGlyValThrGlyValTrrArgSerAspCysGly 690
 DB 2098 ATCTGGGACCGACGACCAAGCTGGAACGACATCGGCTGTACCGCTCCGACGACACGCG 2157
 QY 691 ThrThrTrrPValLeuIleAsnAspAspGlnHisGlnTrrGlyAsnTrrPLeuAlaIle 710
 DB 2158 AGCACCTGGACGCGGTCAATGACAGGACCACTACTTCGGGC---CCACCATGATC 2214
 QY 711 ThrGlyAspHisAlaAsnLeuArgValTrrIleGlyThrAsnGlyArgGlyIleVal 730
 DB 2215 GAGGCGGACCCCAAGTCTTACGCGCGGTGTATCTAGGCACGAACGCGCGGTATCGTG 2274
 QY 731 TyrGlyAspIle 734
 DB 2275 TACGCGGACCTT 2286
 RESULT 10
 ADH19117
 ID ADH19117 standard; cDNA; 2646 BP.
 AC ADH19117;
 XX 11-MAR-2004 (first entry)
 DE Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme precursor cDNA.
 XX xyloglucan oligosaccharide degradation; genetic engineering; ss; gene;
 KW precursor.
 XX Geotrichum sp. M128.
 OS
 XX Location/Qualifiers
 FH Key 120..2558
 CDS /tag= a
 FT /product= "Geotrichum sp M128 xyloglucan oligosaccharide-
 FT degrading enzyme precursor protein"
 XX
 PN EF1350844-A2.
 XX
 XX 08-OCT-2003.
 XX 25-MAR-2003; 2003EP-00251866.
 XX 25-MAR-2002; 2002JP-00083433.
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX Yaoi K, Mitsuishi Y;
 XX

DR WPI; 2004-100948/11.
 XX P-PSDB; ADH19118.
 PT Novel xyloglucan oligosaccharide-degrading enzyme with a different
 XX degradation mechanism from known enzymes.
 PS Claim 11; SEQ ID NO 11; 39pp; English.
 XX
 CC The invention relates to a novel xyloglucan oligosaccharide-degrading
 CC enzyme with a different degradation mechanism from known enzymes. The
 CC primary structure and polynucleotide structure of the xyloglucan
 CC oligosaccharide-degrading enzyme provided by the present invention allows
 CC high-purity polypeptides having a xyloglucan oligosaccharide-degradation
 CC activity to be prepared at a low cost through a genetic engineering
 CC process. The current sequence is that of the *Geotrichum* sp. M128
 CC xyloglucan oligosaccharide-degrading enzyme precursor cDNA of the
 CC invention.
 XX
 SQ Sequence 2646 BP; 550 A; 866 C; 787 G; 443 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 7.38e-63 Length: 2646
 Score: 1117.00 Matches: 281
 Percent Similarity: 50.37% Conservative: 124
 Best Local Similarity: 34.95% Mismatches: 279
 Query Match: 27.68% Indels: 120
 DB: 12 Gaps: 31

US-09-917-376-3 (1-740) x ADH19117 (1-2646)

QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyPheValAspGlyValPheAsn 25
 DB 198 TACGAGTTCAGATGTCGATCGCGCGCGGGTACATTACCGGATTTGTCGCGAC 257
 QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
 DB 258 CCAAGACCAAGGACCTGCTAGCGCGCGACGACATTTGGCGCGCTACCGTGGGAC 317
 QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyr 65
 DB 318 GCAGGACAGTCCAGTGGATCCGCTCAACACTTTATCGAGGCGCGAGCATGAACATT 377
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
 DB 378 ATGGGACCGAGTGCATCGCTGGACCCCAACACCCCGACAGCTGTACTCGCGAC 437
 QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104
 DB 438 GGGCGCTATGTCGCGACAGTGG-----GGGCGTTCTATGTGTCGGAAGAC 485
 QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124
 DB 486 CGCGGCCAGTCGTTTACAACTACGAGTCGCGCTTCCGATGGGCGCCCAACGACATGGGA 545
 QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144
 DB 546 CGCAACATGCGAGCGCTCGCTGCTCAACCGCTTCACTCGAACGAGGTCTGGATGGGT 605
 QY 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
 DB 606 ACGGTACA---GAGGGTATCTGGAAGATTCGGACCGCGCCCAAGACCTGGACAAACGTC 662
 QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGln 184
 DB 663 ACGTCCATCCCGGACCGCTTC-----ACCAACGGTATCGGATACACG 704
 QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204
 DB 705 TCG-----GTCATTTTCGACCCC-----GAA 725
 QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224
 DB 726 CGTAAATGGCACCATCTACGCGGCGGACTGCCCGCGAGGGC---ATGATGTCACGCA 782

QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241
 DB 783 GACGCGGTGTCGTCTGTCGGAGCCAGTGGCGGCGCCAGCGCTCCAGCTGGTGGTCAACAGGACC 842
 QY 242 -----ProHisLysGlyVal 246
 DB 843 ACGGCGGTTCCTCCGCAAGAAGCCCGTCGATCGCGCCGACGCCATGAAGTCGT 902
 QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266
 DB 903 CTCACCCCC-----AACTTCTACGTGACTTAGCGGACTACCTCGTCAATGGGC 956
 QY 267 GlySerSerGlyAspValTrpLysPheSerValTrpSerGlyThrTrpThrArgIle--- 285
 DB 957 GTCAGTTCGGGAGTCTGGCGCCAGAACCGACCTCGGCGCGCTGGGACGACATTACT 1016
 QY 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
 DB 1017 CCCGCGTCGCGCAACTCGTCGCTGCCGCTACAAACACACGACGCTTCCCTCGGCGCGA 1076
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
 DB 1077 TTTTGGGTCTCAGCGTCGACGCGACCAACCCCAACCGCTCTGTCGTATCACC---CTC 1133
 QY 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336
 DB 1134 GACCGGACCCCGGACCCGCTCGACGACATCTACCTCTCAACGATCCGCGCGGAC 1193
 QY 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347
 DB 1194 TGGAGGACGTCACCCAGCTCTCGTCCCGCTCAACCTCGAAGTAACTGGGCGCACCCG 1253
 QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpIleThrPheGlyVal 367
 DB 1254 ACTAACCGCGCGGTAC---AAGGACGCGACGCTGTTCCGTGGCTCGACTTCAACAAC 1310
 QY 368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379
 DB 1311 GGTCCCAAGTGGGGGGATACGGTCGCGCGACGCTACGCCGCGCTCAACCAAGTTGGC 1370
 QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399
 DB 1371 TGGTGGATGAGCGCTGTCTTATCGATCGTTCAACCCCGACGACCTGTATGTACGCGACG 1430
 QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 415
 DB 1431 GGGGCGACCATCTGGGCGACCGACACGCTCTCCGTCGAGAGGACTGG----- 1481
 QY 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla 431
 DB 1482 -----GGCGCGAGCTGGTACCTCCAGATCGACGCTATCGAGGAGATGCG 1526
 QY 432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451
 DB 1527 ATCTGTGCTCGCTCGCCCAAGAGCGCGGCGCTCTCTGTGGGCATCGGTGACATT 1586
 QY 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471
 DB 1587 AGCGCATGAAGCACGACGACCTCCCAAG---CCCCAGAAGATGTTTGTGCGGCCCG 1643
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491
 DB 1644 TTCTCAACCTCGACGATCGACGCTCGGGCAACTTCCCAACGTTGTGTCGCGGCC 1703
 QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
 DB 1704 GGATCTCGGACACGACGACGCGCTGCGCGGCTGCGTACGCGACTGACGGC 1763
 QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThr----- 526
 DB 1764 GGAGACGCTGAGCCATCTTCCCTACCTGCTGCTGCGATGAAACCGGACCACTACCG 1823


```
QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 544
DB 1824 GGCACGAGTTCGATCGACGCGGCGGAGCCGATCGTGTGTGTCACCAAGCTTGAC 1883
QY 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564
DB 1884 GAGCAGGCTCGGACCGTGTACTCGCAGCAGTATGCAAGACGTGTCT-----1934
QY 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578
DB 1935 ---GTTCCGCTGGCGAGCTGAAGGCCGAGACTGCAATGTGCTCTCGCAAGTCCAG 1991
QY 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyVal 598
DB 1992 GATGCGACGTCTACGCTACCGATCGCGGCAAGTTCCTCGTCTCGACGACGCGGGAAG 2051
QY 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
DB 2052 TCGTATGCGCCCAAGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 2105
QY 619 AlaValProGlyLysGluGlyAspLeuTrpLeu-----AlaAlaSerSerGlyLeuTyrHis 637
DB 2106 GTGAACCTCGGTGGCGGCGACGCTCGGTGCTGTTCGCGAGGCGGTCTCTTCCAC 2165
QY 638 SerThrAsnGlyLysSerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656
DB 2166 TCGACCGACTTTCGCGCCTCGTTCACGAGGTAGTACCGCAACGCGACCTCGTGAGC 2225
QY 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672
DB 2226 GTGCGGCGCCCAAGTCCAGTCGCGAGCGCAGAGGCTAGCGCGCCCTCGCGGTCTTC 2285
QY 673 ValValGlyThr-----IleGlyValThrGlyAlaTyrArgSerAspCysGly 690
DB 2286 ATCTGGGCGACCGACAGCCTCGAGCGACATCGGCTGTACCGTCCGACGACACGCG 2345
QY 691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710
DB 2346 AGCACCTGACGCGCGTCAATGACGAGGACGACAACTACTCGGC-----CCCAACCATGATC 2402
QY 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyValGlyIleVal 730
DB 2403 GAGCGCGACCCCAAGGTCTACGGCGCGGTATCTAGGCGCAACGCGCGGTATCTGTG 2462
QY 731 TyrGlyAspIle 734
DB 2463 TAGCGGACCTT 2474

RESULT 11
ID ADH19123
XX ADH19123 standard; DNA; 2481 BP.
AC ADH19123;
XX
DT 11-MAR-2004 (first entry)
DE Geotrichum sp M128 xyloglucan oligosacch-degrading enzyme-derived DNA.
XX xyloglucan oligosaccharide degradation; genetic engineering; ds; gene;
KW His tag.
XX
OS Synthetic.
XX Geotrichum sp. M128.
XX
FH Key Location/Qualifiers
FT CDS 1..2481
FT /tag= a
FT /product= "Geotrichum sp M128 xyloglucan oligosaccharide-
XX degrading enzyme-derived protein"
FN EP1350844-A2.
XX
PD 08-OCT-2003.
```

```
XX 25-MAR-2003; 2003EP-00251866.
XX PF
XX 25-MAR-2002; 2002JP-00083433.
XX PR
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX PA
XX Yaoi K, Mitsuishi Y;
XX PI
XX WPI; 2004-100948/11.
XX DR P-PSDB; ADH19124.
XX
XX Novel xyloglucan oligosaccharide-degrading enzyme with a different
XX degradation mechanism from known enzymes.
XX Example 7; SEQ ID NO 17; 39pp; English.
XX
XX The invention relates to a novel xyloglucan oligosaccharide-degrading
XX enzyme with a different degradation mechanism from known enzymes. The
XX primary structure and polynucleotide structure of the xyloglucan
XX oligosaccharide-degrading enzyme provided by the present invention allows
XX high-purity polypeptides having a xyloglucan oligosaccharide-degradation
XX activity to be prepared at a low cost through a genetic engineering
XX process. The current sequence is that of the Geotrichum sp. M128
XX xyloglucan oligosaccharide-degrading enzyme-derived DNA of the invention
XX which comprises the wild-type enzyme sequence plus a C-terminal region
XX including a Histidine tag originating from vector pET29a(+).
XX
XX Sequence 2481 BP; 496 A; 836 C; 748 G; 401 T; 0 U; 0 Other;
XX
```

```
Alignment Scores:
Pred. No.: 1,07e-62 Length: 2481
Score: 1114.00 Matches: 281
Percent Similarity: 50.25% Conservative: 123
Best Local Similarity: 34.95% Mismatches: 280
Query Match: 27.60% Indels: 120
DB: 12 Gaps: 31
US-09-917-376-3 (1-740) x ADH19123 (1-2481)
```

```
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
DB 13 TAGCAGTTCGAAGATGTCGCGATCGCGCGCGGTACATTACCGGGAATTGTCGCGCAC 72
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
DB 73 CCAAGACCAAGGACCTGTGTACGCGCGCGCACGACATTGCGCGCGGTACCGCTGGAC 132
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
DB 133 GCAGGCACTGCAAGTGGATCCCGCTCAACGACTTATTCAGGCGCGGAGGACATGAACATT 192
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
DB 193 ATCGGCGACCGAGTCGATCGCGCTGGACCCCAACACCCGACAGGCTGTACCTCGCGCAG 252
QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104
DB 253 GGGCGCTATGTCGCGCAGCAGTGG-----GCGGCGTTCCTATGTGTCGGAAGAC 300
QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124
DB 301 CGCGCGCAGTCGTTTCAATCTACGAGTCGCCGTTCCTCGATGGCGGCGCAACGACATGGGA 360
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144
DB 361 CGCAACAATGGCGAGCGCTCGCTGTCACCCGTTCACTCGAACGAGGTCTGGATGGGT 420
QY 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
DB 421 ACGCGTACA--GAGGGTATCTCGAAGAGTTCGACCGCGCAAGACCTGGACAACGTC 477
QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGln 184
```



```

Db 478 ACCTCCATCCCGGACGGCTTC-----ACCAACGGTATCGGATACACG 519
Qy 185 SerAspIleGlnGlyValValTrrpValAlaPheAspLysSerSerSerLeuGlyGln 204
Db 520 TCG-----GTCATTTTCGACCCC-----GAA 540
Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrrpSerArg 224
Db 541 CGTAATGGCACCATCTACGGAGCGGCGACTGCCCGCAGGGC---ATGTACGTACCGCAC 597
Qy 225 AspGlyGlyAlaThrTrrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241
Db 598 GACGGCGGTCTCTCGTGGAGCCAGTGGCGGCCCGCGCTCCAGCTGGCTCAACAGGACC 657
Qy 242 -----ProHisLysGlyVal 246
Db 658 ACGGGCGGTTCCTCCGGACAAGACCGCGCTCGATCGCGCGCGCACGCCCATGAAAGTCGCT 717
Qy 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266
Db 718 CTCACCCCC-----AACTTCCTCTACGTGACTTACCGCCGCACTACCTCTGGTCCATGGGG 771
Qy 267 GlySerSerGlyAspValTrrpLysPheSerValThrSerGlyThrTrrpThrArgIle--- 285
Db 772 GTCAAGTTCGGCAAGTCTGGGCGCAGAACCGCACCTCGGGCGCTGGGACGACATTACT 831
Qy 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
Db 832 CCCGCGTCGGCAACTCGTCGCTGCCCGCTACCAACACAGAGAGCTTCCTCGCGGGCGGA 891
Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
Db 892 TTTTTCGGGTCTACGCTCGACGCGCAGCAACCAACCCGCTCTCGTCGTATCACC---CTC 948
Qy 320 SerTrrpTrrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336
Db 949 GACCGCACCCGACCGCGCTCGACAGCATCTACTCTCAACCGATGCGCGGCGGAC 1008
Qy 337 TrrpThrArgIleTrrpAspTrrpThrSer-----TyrPro 347
Db 1009 TGGAAGGAGGTCACCCAGCTCTCGTCCCGTCCCAACCTCGAAGGTAACTGGGGCCACCG 1068
Qy 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrrpLeuThrPheGlyVal 367
Db 1069 ACTAACCGCGCGCGGTAC---AAGGACGGCACGCTGTTCGTCGTCTCGACTTCAACAAC 1125
Qy 368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379
Db 1126 GGTCCCGAGTGGGGGATACGTCGCGCGCACCGTACGCCCGCGCTCACCAAGTTGGC 1185
Qy 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399
Db 1186 TGGTGGATGACGCGTGTATCGATCGTTCAACCCCGAGCACCTGATGATACGGCACG 1245
Qy 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrrpAspSerGly 415
Db 1246 GGGCGCACCATCTGGCGCGACCGACGCTCTCCGCTGTTCGAGAAGGACTGG----- 1296
Qy 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla 431
Db 1297 -----GGCCCGAGCTGGTACTCTCCAGATCGACGGTATCGAGAGAAATGCG 1341
Qy 432 ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451
Db 1342 ATCTGTGCTCGCTCGCTCCCAAGAGCGCGCGGCTCTCTGTCGGGCATCGGTGACATT 1401
Qy 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471
Db 1402 AGCGGCATGAAGCACGACGACCTCACCAG---CCCCAAGAAGATGTTTGGTGGCGGCCAG 1458
Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491

```

```

Db 1459 TTCTCCAACTCGACGATCGACGCTCGCGGCAACTTCCCAACGTTGTGTCTCGCGCGC 1518
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
Db 1519 GGATCCTCGGACACGAGTAGCAGACGCGCTGCGCGCGGTGCGTACGCGACTGACGCG 1578
Qy 512 GlyLysAsnTrrp-----PheGlnGlySerGluProGlyGlyValThrThr 526
Db 1579 GGAGACGGTGGACCATCTTCCCTACCTTCCCTGCGCATGAACGCGAGCCACTACCA 1638
Qy 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrrpAlaPro-----Gly 544
Db 1639 GGCAGACGATTCAGTCGACGCGAGCGGACGATCGTGTGTCGACCAAGCTTGAC 1698
Qy 545 AspProGlyGlnProValValTrrpAlaValGlyPheGlyAsnSerTrrpAlaAlaSerGln 564
Db 1699 GAGCAGCGCTCGGACCGTGTACTCGACGACTATGSCAAGACGTGTCT----- 1749
Qy 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578
Db 1750 ---GTTCCCGCTGGCGACCTGAAGGCCGAGACTGCCAATGTCTCTCGACCAAGTCCAG 1806
Qy 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyVal 598
Db 1807 GATGGCACGTTCTACGCTACCGATGGCGGCAAGTCTTCTCGTCTCGACGCGCGGAA 1866
Qy 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
Db 1867 TCGTATGCGCCCAAGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 1920
Qy 619 AlaValProGlyLysGlyValAspLeuTrrpLeu---AlaAlaSerSerGlyLeuTyrHis 637
Db 1921 GTGAACCCCTGGTGGCGCGGACGCTGGGTGCTGTCCCGAGGCGGTCTCTTCCAC 1980
Qy 638 SerThrAsnGlySerSerTrrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656
Db 1981 TCGACCGACTTGGCGCTCGTTCAAGGGTAGTACTCCGCCAACCGCACCTCGTGAGC 2040
Qy 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672
Db 2041 GTCGCGCCCCCAAGTCCAGTCGAGCGGCAAGAGGTAGCGGCCCTCCGCGGTCTTC 2100
Qy 673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 690
Db 2101 ATCTGGGACCCGACAGCCTGGAGCGACATCGCCTGTACCGTCCGACGACCAACGC 2160
Qy 691 ThrTrrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrrpGlyGlnAlaIle 710
Db 2161 AGCACCTGGACGCGCTCAATGACGAGGACCAACTACTCGGGC---CCCACCATGATC 2217
Qy 711 ThrGlyAspHisAlaAsnLeuArgValTrrpIleGlyThrAsnGlyArgGlyIleVal 730
Db 2218 GAGGCGGACCCCAAGTCTACGGCGCGTGTATCTAGGCACGACGCGCGGTATCGTG 2277
Qy 731 TyrGlyAspIle 734
Db 2278 TACGCCGACCTT 2289
RESULT 12
ADR90297
ID ADR90297 standard; DNA; 2268 BP.
XX ADR90297;
AC ADR90297;
DT 16-DEC-2004 (first entry)
XX Geotrichum sp. M128 xyloglucanendohydrolase mature DNA.
XX endo-type xyloglucanendohydrolase;
KW xyloglucan-specific endo-b-1,4-glucanase; glucose; xylose; xyloglucan;
KW plant cell differentiation; mature; ds; gene.
XX Geotrichum sp. M128.
OS.

```


XX FH Key Location/Qualifiers
 FT CDS 1..2268
 FT /*tag= a
 FT /partial
 FT /product= "Geotrichum sp. M128 xyloglucanendohydrolase
 FT mature protein"
 FT /note= "No start or stop codon"
 XX JP2004261037-A.
 XX PN
 XX PD 24-SEP-2004.
 XX PF 28-FEB-2003; 2003JP-00053286.
 XX PR 28-FEB-2003; 2003JP-00053286.
 XX PA (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.
 XX WIPI; 2004-665466/65.
 XX P-PSDB; ADR90298.
 XX Novel xyloglucan hydrolase derived from Geotrichum genus, having ability
 PT to hydrolyze beta-1, 4-glycoside linkage of glucose residue in
 PT xyloglucan, useful for elucidating structure or function of xyloglucan.
 XX Claim 7; SEQ ID NO 14; 86pp; Japanese.
 XX The invention relates to a novel xyloglucanendohydrolase (xyloglucan-
 CC specific endo-b-1,4-glucanase) enzyme derived from the Geotrichum genus
 CC and having the ability to hydrolyse the beta-1,4-glycosidic linkage of a
 CC glucose residue, but not that of a xylose residue contained in a
 CC xyloglucan. The polypeptide of the invention may be useful for
 CC elucidating the structure or function of a xyloglucan having an important
 CC role in the differentiation of a plant cell. The current sequence is that
 CC of the Geotrichum sp. M128 xyloglucanendohydrolase mature DNA of the
 CC invention.
 XX Sequence 2268 BP; 413 A; 795 C; 717 G; 343 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 5,24e-54 Length: 2268
 Score: 979.00 Matches: 264
 Percent Similarity: 48.43% Conservative: 122
 Best Local Similarity: 33.12% Mismatches: 279
 Query Match: 24.26% Indels: 132
 DB: 13 Gaps: 33

US-09-917-376-3 (1-740) x ADR90297 (1-2268)

QY 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30
 DB 22 GTGACATTCGGTGGCGGCTTCATCTCGGCTCTGGCACACCGCGAGAGGAC 81
 QY 31 IleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 50
 DB 82 CTGATCTACGACGACAGACATTTGGCGGCACGTACAGGTGGAACCGGCCAAGTGGGAG 141
 QY 51 TrpIleProLeuAspTrpVal-----GlyTrpAsnAsnTrpGlyTyr 65
 DB 142 TGGGAGCCCATCGGACTTTATCATCAACACGCGCTCGCGGCAACCGCGCCACCTTG 201
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
 DB 202 CTGGCACCAGAGGACATTGGCTCGACCCCAACACCGACCGCTGTACCTCGCGCAA 261
 QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
 DB 262 GGGACTATGTG---CAGTGGGACCG---TGGGCGGCTTCTCTCTCGGACGACGCG 315
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
 DB 316 GGCACAGACGTTCAACGAGTAGTCCGCTCGCGCTCCCATGGGCGCCCAACGACATGGGCGCG 375

QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
 DB 376 AATGGCGGCGAGCGCTCGCGGTCAACCGCGACTCGACCGAGCTGTGGTTGGCTCG 435
 QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
 DB 436 CGCAGC---CAGGGCGCTGTGGCGCTCCACTGACCGCGCTCAGACGCGGTGCGCATGAAC 492
 QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGlnSer 185
 DB 493 CAGCTCCCGCAGCTCGACGTAC-----GGATC----- 522
 QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 205
 DB 523 -----GGCATCATCTCGGTCTTCGACCCCAAGAACGTCGCGCAGCATCATGTC 573
 QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 225
 DB 574 GCTTCGCACGCGCTCGCGGCTGTGG-----GTCACGTGG-----GAC 612
 QY 226 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThr----- 238
 DB 613 GCGCGCGCAACTGCTGCCAGGTGGCGCGCCACTACGCGAGTGTGTCAGACTGGACAAAG 672
 QY 239 -----GlyPheIleProHisLysGly 245
 DB 673 AGCATTTGCTCGCTCGCGCACCGCATCCAGAGCTCGGGGCCCCCTCCCGATCAAGATC 732
 QY 246 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 265
 DB 733 GCACTGGGCAAG---AACGCGAGCTGTACATTACCTACTCGGACGACCGGCGCCCTGG 789
 QY 266 AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle 285
 DB 790 GCGCTCTCTACGCGCGAGGTCTGCTCTACGATCCCAACCAACGCAACTGGAAGCACATC 849
 QY 286 SerProValProSerThrAspThrAlaAsnAspTyr----- 297
 DB 850 ACC-----CCCTCGCGCGAGGGCGCCACACATGTTACCTCCCGCCACTGGCAACAAAG 903
 QY 298 -----PheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 314
 DB 904 GTCGTTCCCGGAGGCTGGAAAGGTATCTCGTGGT-----AACGGGACACCGCTCGTC 957
 QY 315 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 334
 DB 958 GTGTGACGCTCGACGCAACGCGCAAGACTCA---GTGTACCTCTCTCGCGCGCGCGC 1014
 QY 335 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 354
 DB 1015 AACTGTGGAAGACCTCGGCAAGCTGACCAACCGCCCGCGCGCGCGCGGCACTCGCAG 1074
 QY 355 LeuAspIleSerAlaGlu-----ProTrpLeuThrPheGlyVal 367
 DB 1075 AAGAGTGGACGCCAAGCTGCGCAACGCGCACTCGGCTCCCTCGCTCTCGTTCAGAAC 1134
 QY 368 GlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIle 387
 DB 1135 CGCGGCTCGGTATTGTG-----GGATTGCTGGTGGCTCGCGCGCATCTCTCTC 1185
 QY 388 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 407
 DB 1186 GACCCCTTC---TCGGACCGCTCTGTACGCGCAGCGCTGTGATCTGGGCGGACTGAC 1242
 QY 408 AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMet----- 423
 DB 1243 GCGCTGTGCGCGCGCGGACTCG-----AACCGAGCGCGAGCTGGTGTACATCAAC 1290
 QY 424 ValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGly---Ala 442
 DB 1291 ACAGAGGCGCATCGAGGAGCGGCGATCTGTGTCTCTCAAGTCGCGCGCTCGCGCGCGCG 1350

Qy	368	GlnProAsnProProValProSerProLysLeuGlyThrMetAspGluAlaMetAlaIle	387
Db	1396	CGGGCTCGGATTGTC-----GGATTCCGCTGGTGGCTCCGCCCATCTCCTCCTC	1446
Qy	388	AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaMetLeuTyrAlaThrAsn	407
Db	1447	GACCCCTTC---TCGGACCGCTCCTGTATCGGCACAGGCGCTGTGTATCTGGGCGACTGAC	1503
Qy	408	AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMet-----	423
Db	1504	CGCGTGTCCGCCCGCACTCG-----AACCAGCGCGCGAGCTGGTATACATCAAC	1551
Qy	424	ValLysGlyLeuGlnGluThrAlaValAsnAspLeuIleSerProProSerGly---Ala	442
Db	1552	ACAGAGGCGCATCAGAGAGCGCGATCTCGTCTCAAGTCGCGCGCTCGGGGCCCGCG	1611
Qy	443	ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal	462
Db	1612	CACCTCTTCAGCGGCATGTACGACCTCGGGCATCGGCCATGATGACTTC---TCGCTC	1668
Qy	463	ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGlu	482
Db	1669	CCGCAGCCCATGTACAGCAAGCCGACCTTCAGCAGCAGCAGCGCCTCGACTTTGGGGC	1728
Qy	483	LeuAsnProSerIleIleValArgAlaGlySerPheAsp-----ProSerSerGlnPro	500
Db	1729	CGCGCGCCCAATGTGTCCGCGCTCGGGCGCAACGACCATCTCGACGGCGGCTCGCG	1788
Qy	501	AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp-----PheGlnGly	518
Db	1789	GGGTGCACGCGGGCGGTACACGACCAACAGCGGCGACAGCTGGAGCTCTTTCAGACC	1848
Qy	519	SerGluProGly---GlyValThrThrGlyThrValAlaIleSerAlaAspGlySer	537
Db	1849	TGTTGTCGCCAGCCTCGAGGTCCGTAAACGCGCGCACGATCGCGTCGCGCGGCGCAAG	1908
Qy	538	ArgPheValTrpAlaProGlyAsp---ProGlyGlnProValValTyrAlaValGlyPhe	556
Db	1909	ACGTTTGTGTGTGCCCCAGCAAGCGGCGACGCGGCGCCGTACATCATCGGCGACTAC	1968
Qy	557	GlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln---IleArgSerAsp	575
Db	1969	GGCAAGACGTGGACTGCCCGCGCGGCTGAGCAACGACGACGAGTATCGCGCGCAC	2028
Qy	576	ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp	595
Db	2029	CGCGTCGAGCCCAACACATTCTACGTGTACGTTCGAGGGCGACTTCTCGTCTCGACCGAT	2088
Qy	596	GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal	615
Db	2089	GGCGGCAAGAGCTACACCAAGAGGGCAACGGCCCTCCCGTCTCGACGCG-----	2139
Qy	616	MetPheHisAlaValProGly-----LysGluGlyAspLeuTrpLeuAlaAla	631
Db	2140	---TACACGGGCGACCCCGTTCACGAGCAACTGCGTCCGGCGGAGCTGTGGTCTCCGTC	2196
Qy	632	SerSer---GlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGly	650
Db	2197	AAGGGCGTTGGCATCTACCAAGCACCGACTTTTGGCAACACATTTCACCGCGCTTCGGGC	2256
Qy	651	ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAla	670
Db	2257	TCGGGCTCCAGTCTCAAT-----CCCGCT-----CCCGCT	2280
Qy	671	ValPheValValGlyThr-----IleGlyGly	679
Db	2281	GTGTTTCAGCATCGCGCGCGCCGACAGCCCAACGCCACGAGACGCTCTTCTCTGGGGC	2340
Qy	680	Val-----ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrp	693
Db	2341	ATCCCGTCGGCGTCGACCGCGGCGCTGTATATGTTCGACCGCAACAGCGCGGTTGTGG	2400

Qy	694	ValleulleAsnAepAepGlnHisInTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp	713
		:::	
Db	2401	ACGGCCTCAACGACGACGCGCAACATATGTTGGC---GCCACCGTGATCAGCGGCAT	2457
Qy	714	HisAlaAsnLeuArgValTyrIleGlyThrAsnGlyVArgGlyIleVal	730
Db	2458	CCCGCATCTACGGCCGCTTACATGGCATGAACGCGCGGGGCATCATC	2508
RESULT 15			
ADD42054			
ID	ADD42054	standard; DNA; 5698 BP.	
XX	AC		
XX	AC	ADD42054;	
XX	AC		
DT	15-JAN-2004	(first entry)	
XX	XX	Trichoderma reesei Family 74 xyloglucanase-encoding DNA, SEQ ID NO:1 #1.	
DE	XX	Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage	
KW	KW	xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;	
KW	KW	ethanol production; detergent composition; fabric treatment;	
KW	KW	textile treatment; ds.	
XX	XX		
OS	XX	Hypocrea jecorina.	
XX	XX		
PN	PN	WO2003089598-A2.	
XX	XX		
PD	30-OCT-2003.		
XX	XX		
PF	17-APR-2003; 2003WO-US011831.		
XX	XX		
PR	19-APR-2002; 2002US-0373987P.		
XX	XX		
PA	(NOVO) NOVOZYMES BIOTECH INC.		
XX	XX		
PI	Michael R, Zaretsky E, Haas J;		
XX	XX		
DR	WPI; 2003-845528/78.		
DR	P-PSDB; ADD42055.		
XX	XX		
PT	New polypeptides having Family 74 xyloglucanase activity, and encoding		
PT	nucleic acid molecules, useful for degrading cellulose- and hemicellulose		
PT	-containing biomass to ethanol or as a detergent.		
XX	XX		
PS	Claim 1; SEQ ID NO 1; 96pp; English.		
XX	XX		
CC	The invention relates to a Family 74 xyloglucanase (ADD42055) from the		
CC	fungus <i>Trichoderma reesei</i> (<i>Hypocrea jecorina</i>), and nucleic acids encoding		
CC	it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic		
CC	linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.		
CC	The invention also relates polypeptide sequences at least 70% identical		
CC	to the enzyme, expression vectors and host cells comprising a nucleic		
CC	acid of the invention, the recombinant production of the enzyme, and		
CC	mutant enzymes and the nucleic acids encoding them. The xyloglucanase of		
CC	the invention can be used in the degradation of cellulose- and		
CC	hemicellulose-containing biomass to produce ethanol. It can also be used		
CC	in a detergent composition for treating fabric during a machine washing		
CC	cycle. The present sequence represents a <i>Trichoderma reesei</i> Family 74		
CC	xyloglucanase-encoding DNA. Note: The present sequence (referred to as		
CC	SEQ ID NO:1) is given as shown in the sequence listing. A comparison of		
CC	this sequence with that shown in figure 1 (also referred to as SEQ ID		
CC	NO:1) shows that the present sequence is twice as long as the figure 1		
CC	sequence (ADD42060) and contains alternate 60 bp blocks of sense strand		
CC	and complementary strand sequence.		
XX	XX		
SQ	Sequence 5698 BP; 1145 A; 1704 C; 1704 G; 1145 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	4, 79e-47	Length:	5698
Score:	878.50	Matches:	334
Percent Similarity:	29.9%	Conservative:	119
Best Local Similarity:	22.0%	Mismatches:	270
Query Match:	21.7%	Indels:	793

DB:	10	52	Gaps:
US-09-917-376-3 (1-740) x ADD42054 (1-5698)			
Qy	7	ThrTrpSerAsnValAlaIle---	GlyGlyGlyGlyPheValAspGlyIleValPheAsn 25
Db	242	TCATGAAGAAACGTCAAGCTTCGGCGCGCGCGCTTCGTCCCGGCATCATCTCCAA	301
Qy	26	-----GluGlyAlaPro-	29
Db	302	AGTACCTCTTCGAGTTCGAGCCGCGCGCCGCGAAGCAGGGCGGTAGTAGAAGTT	361
Qy	30	-----GlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrrp----	44
Db	362	CCCAAGACAAAAGCGGTAGCATATGACGAACAGATATTGGCGGGCTGTACCCGCTCAAA	421
Qy	45	-----AspAla 46	
Db	422	GGGTCTGTTTTCCGCAACGTATACGTGCTTGTCTATACCGCGCGACATGGCGGAGTTC	481
Qy	47	AlaAsnGlyArgTrrpIleProLeuLeuAspTrrpVal----	GlyTrrp----- 60
Db	482	GCCGACGATCATGGACCGCGCTCACGGATGGGATTGCTGATATGCGGCTGSCACAAG	541
Qy	61	-----AsnAsnTrrpGly-----	Tyr 65
Db	542	CGGCTGTGAGTACCTGGCGGCAGTGCCTACCTTAACGACTATTACGGCCGACCGGTGTT	601
Qy	66	AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIysValTrrpAlaIaVal	85
Db	602	TGGGGCATACGCTGTTCGCTGTGATTCGCGACGACGATCAAAAGGTGTATGCCGCGATG	661
Qy	85	-----	85
Db	662	ACCCGCTAGCTGGACACACCGCACTAGGCGTCTCTGTAGTCTTTCCACATACCGCGTCC	721
Qy	86	GlyMetTyrThrAsnSerTrrpAsp-----	93
Db	722	GGCATGTATACGAACAGCTGGTC- TGTGATGTCTCTCAGATCTAGACCTATGATTGGACG	780
Qy	93	-----	93
Db	781	GCCGTACATATGCTTGTTCGACACACACTACAGAGTCTAGATCTGGATCTAACTCTGC	840
Qy	94	-----ProAsnAspGlyAlaIleLeuArgSerSerAspGln 105	
Db	841	CTGACATTGGCCATATATAGGATCCGAGTAAAGGACCATCATTCCTCTCTCAGACCGC	900
Qy	105	-----	105
Db	901	GACTGTAAACCGGTATATATCCCTAGGCTCATTAACCTCGGTAGTAAAGCAGCATCTCGCG	960
Qy	106	GlyAlaThrTrrpGlnIleThrProLeuProPheIysLeuGlyAsnMetProGlyArg	125
Db	961	GGCGCAACGTGGTCTTCCACCAACTTGCCTTCAAAGTCGGGGGTAAACATGCCAGGACGC	1020
Qy	125	-----	125
Db	1021	CCGGTTGCACAGGAAGTGGTGAACCGGAAGTTTCAGCCCCCATTTGTAGGTCCTCGCG	1080
Qy	126	GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla	145
Db	1081	GGAGCCGAGAGCGTCTGGCTGTCGATCCGGCCCAACTCCCAACATCATCTACTTTGGTGCT	1140
Qy	146	Pro-----	146
Db	1141	CCTCGGCTCTTCGACAGCCAGCAGCTAGGCGGTTGAGGTTGTAGTAGATGAACACCA	1200
Qy	147	---SerGlyIysGlyLeuTrrpArgSerThrAspSerGlyAlaThrTrrpSerGlnMet---	164
Db	1201	CGCTCAGGAACCGCCCTCTGGAAGTCTACGACCGGCGGCTGACCTTTTCCAAAGTCTCG	1260
Qy	165	-----Thr 165	

Db 2341 AAAGCAGCTGGCTGAGACC--CTGTTGTACCTCGGGCTAGACCCGCGACCCGCTCGATAGG 2399
QY 362 -----TTP 362
Db 2400 CACTGAGACCTATTACTACAGCATCTCAGTGAGTCACTCTTAAAGATCCGATCGGAATGG 2459
QY 363 LeuThrPhe-----GlyValGlnProAsn----- 370
Db 2460 CTGACTCTGGGATAATGATGTCTGATAGTCACTCAGTGAGAAATGTGAGGTACGCTTACC 2519
QY 371 -----ProValPro----- 374
Db 2520 GACTGACTCTGCTCAAGACTCCCAAGCACCGTGGATCAAGAACAACTTTATCATGATGTG 2579
QY 374 ----- 374
Db 2580 ATGACTGAGACGAAGTTCTGAGGGTTCTGTGGCACTAGTTCTTGTGAAATAGCTACAC 2639
QY 375 -----SerPro-----LysLeuGlyTyrMetAspGluAlaMet 385
Db 2640 TCGACGAGTCAACCGTCCGATCGATGCTCATCAAGCGCCTCGGCTGGATGATGAGTCTCTC 2699
QY 386 Ala----- 386
Db 2700 GGCTCGCTCAGTGGCAGGCTACCAAGTAGTTTCGGGAGCGCCACTTAACTCAGAGAG 2759
QY 387 ---IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405
Db 2760 CAGATTGACCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAACTTTGGC 2819
QY 406 Thr----- 406
Db 2820 GTCTAACTGGTGTGCTGCTGTGTGTGACCGAGATGCGCGTTCCTACTGTTAGAAACGG 2879
QY 407 ---AenAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425
Db 2880 CCGCAGCATCTCACCACCTGGGACAGCGCCACATGTGTCAATCCATCACTGCGCAGAC 2939
QY 426 Gly----- 426
Db 2940 GCGTGTCTAGTGTGTGACCTGTGCGGGTGTACACAGTTAGTTAGTACCGCTCTG 2999
QY 427 ---LeuGluGluThrAlaValAenAspLeuIleSerProProSerGlyAlaProLeuIle 445
Db 3000 CGCATCGAGGAATCTCTCGTCCAGGACCTGGCTCTGCAACCGCGGAAGCAGGATATTG 3059
QY 446 Ser----- 446
Db 3060 GCGTAGCTCTTAAGAGGAGGTCTGGACCGGAGACGTGGGCGCGCTTCGCTCGATATAC 3119
QY 447 ---AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal----- 462
Db 3120 CCGGAGTCCGAGACGACAAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGACATCG 3179
QY 462 ----- 462
Db 3180 CCGCGTACGCTCTGCTGTGTCGAAGTGGAAACGGTCTGTTTGTGGAGCCCTGTAGC 3239
QY 463 ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla--- 481
Db 3240 GCGCAGCGGTCTGGGCAACGCCCATATGGGCCACCTCGACGAGCGTCACTACGCCGG 3299
QY 482 -----Glu 482
Db 3300 AGGTCTGCCACACCGTTGGGTGTACCCGGTGGAGTGTCTCGCAGTGTGATGCGGCC 3359
QY 483 LeuAsnProSerIleIleValArgAlaGlySer----- 493
Db 3360 TACTCGTCAAGAGCGTCTCGCGTCGCGCAACACCGCGCGCACCAACAGAGTGGCCATC 3419
QY 494 -----PheAspProSerSerGlnProAsnAspArgHisValAlaPhe----- 507
Db 3420 TTGAGCCAGTCTCTCGCAGCAGCGCGCGCTGTGTGGCGCGCGTGTGTTCACCGGTAG 3479

QY 508 SerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGly 527
Db 3480 AGTCCGACGGGGCGCGAGCTGGAGCATCGACTACGCGCGCCACACGTCATGAACGGC 3539
QY 528 Gly----- 528
Db 3540 GCGAGCTGCCCGCCGCTGTCACCTCGTAGCTGATGCGCGGCTGTGCAGGTACTTCCG 3599
QY 529 ---ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGly 547
Db 3600 CGCAGGTGGGCTATTTCGGCGCGACGCGCACCATCTCTGGTCGACCGCTCGTCCGGC 3659
QY 548 -----GlnPro 549
Db 3660 GGTGCCACCGATAGCGCGCTGCCCTGTGTAGGAGACAGCTGGCGGAGCAGCGC 3719
QY 550 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsn 569
Db 3720 CTCGACGCTCG--CAGTTCCAGGGCAGCTTTGCTCTCGAGCTCGCGCGGC 3776
QY 570 AlaGln----- 571
Db 3777 GCGCAGCTCGCAGCGCTCAAGTCCGTCGAAACGGAGGACAGCTCGGACGGCGCGC 3836
QY 572 ---IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThr 589
Db 3837 CGGCTCATCGCTCGGACAGAACAGACACAGCGTCTTCTACGCGGCTCCGGATCGACC 3896
QY 590 Phe----- 590
Db 3897 TTTTAGTAGCGGAGCGCTGTTCTTCTGTTGTCGAGAAGATGCGGCGGAGCTAGCTGG 3956
QY 591 ---TyrArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSer 609
Db 3957 AAAAAAGCTCAGCAAGACACCGCGCAGCAGCTTC-----ACGCGCGG--CCCAAG 4004
QY 610 SerGlyAlaValGlyValMet----- 616
Db 4005 TGGGCGAGCGAGGATGAGTGTCTGTCGCGCTGTCGAGTGGCGCGCGCGGTTTC 4064
QY 617 ---PheHisAlaValProGlyLysGluGlyAspLeu 627
Db 4065 GACCGTTCGGTCCCTCGATCCGGATATCGCTGCTCACCGACACCGCGGCGCAGCTTG 4124
QY 628 TrpLeuAlaAlaSerSer----- 633
Db 4125 TATGCTCGACCGACGCGTAGGCCCTATAGCAGAGTGGGCTGTGGCGCGCCGTGCAAC 4184
QY 634 ---GlyLeuTyrHisSerThrAsnGlyGlySerSer----- 644
Db 4185 ATACAGAGTGGGCTGTCTCGGCATATTCCGCTCCACAGACTCGGGCAGACCTTTGGCCAA 4244
QY 644 ----- 644
Db 4245 GTCTCCACCGCCCTGAAGCCGTATAGCGAGGTGTCTAGCGCCGTGTGGAACCGGTT 4304
QY 645 ---TrpSerAlaIleThrGlyValSerSerAlaValAenValGlyPheGlyLysSer 662
Db 4305 CAGAGTGGCGGAGCTCCAAACACTACAGATCGCCCTGGGTGGGCTCAGGCTCGAAC 4364
QY 663 -----AlaProGlySerSerTyr 668
Db 4365 TGGAACTGTATGCTGTGGTTGTGATGTTCTAGCGGAGCCACACCGGAGTCCGAGCTTG 4424
QY 669 ProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla-----TyrArgSer 686
Db 4425 ACCTTGGACATACGATCGGCACC-----GGCCCGTCAGGGGCTCGCTCTACGCCAGT 4478
QY 687 AspAspCysGlyThrThrTrpVal----- 694
Db 4479 GGAGACAGCGCGCCCTCTCTGGAAGCCGTGGCGCGGAGTCCCGAGCGGAGATGCGGTCA 4538


```
QY 695 -----LeulleAsnAspAspGlnHisGlnTyrGly----- 704
Db 4539 CCTCTGTCGCGCGGAGGACCTCGACATCCAGGGCTCCAGGGCTTCGGGCTCCATCGAC 4598
QY 704 ----- 704
Db 4599 AGCACCAAGGTTCGCGCGCAGCGGCTGTAGGTCCCGAGGGTCCCGAAGCCGAGGTAGCTG 4658
QY 705 AsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThr 724
Db 4659 TCGTGTGTTCCAGCGCGCGTCCGCGCAGCACCGCCCGG-----CAAGTCTACGTGGGCACC 4712
QY 725 AsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro 738
Db 4713 AACGGCCGGGGCGTCTTTTACGCTCAGGCGTCTGTGGGGGCC 4754
```

Search completed: October 5, 2005, 09:20:44
Job time : 1225.43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 5, 2005, 07:43:21 ; Search time 7115.69 seconds
(without alignments)
3958.515 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRIGVYDIGGAPSG 740

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool/US09917376/runat_04102005.164328.26911/app.query.fasta_1.2524
-DB=EST_QFMT=fastap SUPFIX=p2n.rst -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09917376 @CN 1 1 12009 @runat_04102005.164328.26911 -NCPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hcc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Match	Length	Description
1	615.5	15.3	704	CF872104 tric029xo
2	615.5	15.3	761	CB902332 tric029xo
3	587.5	14.6	747	CF876916 tric074xd
4	587.5	14.6	814	CB905388 tric074xd
5	415.5	10.3	707	CF880713 tric082xn
6	415.5	10.3	782	CB907625 tric082xn
7	385.5	9.6	693	CF882065 tric029xo
8	361.5	9.0	751	CF868882 tric016xm
9	361.5	9.0	803	CB898982 tric016xm

10	338	8.4	738	7	CF866388
11	338	8.4	794	6	CB896354
12	314.5	7.8	546	2	BF072664
13	306.5	7.6	389	5	BU639045
14	286.5	7.1	713	7	CF875815
15	286.5	7.1	782	6	CB904767
16	282.5	7.0	719	7	CF881775
17	282.5	7.0	775	6	CB908435
18	278.5	6.9	929	7	CF885920
19	278.5	6.9	796	7	CF391910
20	222	5.5	3057	3	AK044947
21	180	4.5	4355	5	AY416870
22	171.5	4.2	296	5	BU641770
23	170	4.2	4631	3	AK082944
24	166.5	4.1	338	2	AW064456
25	153.5	3.8	3623	3	AK083848
26	153.5	3.8	3751	3	AK080277
27	153.5	3.8	4101	9	AY402312
28	151	3.7	4654	3	AK086725
29	150.5	3.7	5041	3	HS080611
30	148.5	3.7	1178	8	CC203127
31	140	3.5	3782	3	AK041860
32	139.5	3.5	5367	9	AK084803
33	138.5	3.4	5370	9	CL976670
34	138	3.4	2847	9	CL982215
35	138	3.4	5073	9	AY407742
36	137.5	3.4	1886	3	AK009469
37	137	3.4	1581	3	BC029959
38	136.5	3.4	1569	9	AY407967
39	136.5	3.4	3200	3	HS080252
40	136	3.4	1674	9	AY412937
41	134	3.3	426	5	BO751758
42	133.5	3.3	995	6	CD780482
43	133.5	3.3	2545	6	CD014011
44	132.5	3.3	769	6	CD788674
45	132.5	3.3	3006	9	CL982143

ALIGNMENTS

RESULT 1
LOCUS CF872104 704 bp mRNA linear EST 31-OCT-2003
DEFINITION tric029xo13.b1 T.reesei mycelial culture, Version 6 October 2003
ACCESSION CF872104
VERSION CF872104.1 GI:38126786
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE 1 (bases 1 to 704)
AUTHORS Diener S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D., Teunissen, P., J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K., Ward, M. and Dean, R.A.
TITLE Characterization of the protein processing and secretion pathways in a comprehensive set of expressed sequence tags from Trichoderma reesei

JOURNAL FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph_dean@ncsu.edu
Seq primer: Lt-F1 primer.
Location/Qualifiers
1. .704
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"


```
/db xref="taxon:51453"
/clone="trico29xol3"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/notes="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Alignment Scores: 6.93e-46 Length: 704
Pred. No.: 615.50 Matches: 118
Score: 615.50 Matches: 118
Percent Similarity: 75.61% Conservative: 37
Best Local Similarity: 57.56% Mismatches: 43
Query Match: 15.25% Indels: 9
DB: 7 Gaps: 4

US-09-917-376-3 (1-740) x CP872104 (1-704)

Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24
Db 98 TTTTCATGGAAGAACGCTCAAGCTCGCGCGCGCGCGCTTCGTCGCCGCGCATCATCTTC 157
Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArg--- 43
Db 158 CATCCCAAGCAAAAGCGGTAGCATATGCACGACAGATATTTGGCGGCTGTACCGCCTC 217
Qy 44 TrpAspAlaAlaAsnGlyArgTyrIlePro-----LeuLeuAspTyrValGlyTyrAsn 61
Db 218 TTCTCCGCGACTCATGACCGCC-GTCACGGATGGATTCGTATATGTCGCGGTGGCAC 276
Qy 62 AsnTrpGlyTyrAsnGlyValValSerIleAlaAspProIleAsnThrAsnLysVal 81
Db 277 AACTGG-----GGCATCGACGCTGTTCGCTTGATCCGACGACGATCAAAAGGTG 327
Qy 82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyValAlaIleLeuArg 101
Db 328 TATCCCGAGTCGGCATGTATACGACAGCTGGGATCCGATATGAGTATGAGCATCATTCGCG 387
Qy 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyLysAsn 121
Db 388 TCGTCAGACCGCGGCGCAACGCTGCTTCACCACTTCGCCCTTCAAAAGTCGGGGTAAC 447
Qy 122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 141
Db 448 ATGCCAGACCGGAGCGGAGAGCGTCTGGTGTGTCGATCCGCGGCAATCCCAACATCATC 507
Qy 142 TyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyValAlaThrTrp 161
Db 508 TACTTTGGTGTCTCGCTCAGGAACAGCGCTCTGGAAGTCTACGACGCGCGGTGACCTTT 567
Qy 162 SerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 181
Db 568 TCCAAGTCTCGCTTCACGCAACTGGGAGTACATCCAGACCCCGAGTATTCACAC 627
Qy 182 GlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSer 201
Db 628 GGCTACCAACGCGATCAGAGGACTCATGTGGGTACGTTTCGATC-ACCAGCAGCAGC 686
Qy 202 LeuGlyGlnAlaSer 206
Db 687 ACCGGNGAGCCACG 701

RESULT 2
CB902332
LOCUS CB902332
DEFINITION trico29xol3 T.reesei mycelial culture, Version 3 april Hypocrea
jecorina cDNA clone trico29xol3, mRNA sequence.
ACCESSION CB902332
VERSION CB902332.1 GI:30116990
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
```

ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE 1 (bases 1 to 761)
Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)

JOURNAL MEDLINE
22803314
PUBMED 12788920
COMMENT
Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.

FEATURES
Location/Qualifiers
source
1..761
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico29xol3"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Alignment Scores: 7.81e-46 Length: 761
Pred. No.: 615.50 Matches: 118
Score: 615.50 Matches: 118
Percent Similarity: 75.61% Conservative: 37
Best Local Similarity: 57.56% Mismatches: 43
Query Match: 15.25% Indels: 9
DB: 6 Gaps: 4
US-09-917-376-3 (1-740) x CB902332 (1-761)
Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24
Db 155 TTTTCATGGAAGAACGCTCAAGCTCGCGCGCGCGCTTCGTCGCCGCGCATCATCTTC 214
Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArg--- 43
Db 215 CATCCCAAGCAAAAGCGGTAGCATATGCACGACAGATATTTGGCGGCTGTACCGCCTC 274
Qy 44 TrpAspAlaAlaAsnGlyArgTyrIlePro-----LeuLeuAspTyrValGlyTyrAsn 61
Db 275 TTCTCCGCGACTCATGACCGCC-GTCACGGATGGATTCGTATATGTCGCGCTGGCAC 333
Qy 62 AsnTrpGlyTyrAsnGlyValValSerIleAlaAspProIleAsnThrAsnLysVal 81
Db 334 AACTGG-----GGCATCGACGCTGTTCGCTTCGTCGCGGACGATCAAAAGGTG 384
Qy 82 TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg 101
Db 385 TATCCCGCAGTCGGCATGTATACGACAGCTGGATCCGATCCAGTATATGAGCATCATCTCGC 444
Qy 102 SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyLysAsn 121
Db 445 TCGTCAGACCGCGCGCAACGCTGCTTCACCAACTTCGCCCTTCAAAGTCGGGGTAAC 504
Qy 122 MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 141
Db 505 ATGCCAGACCGGAGCGGAGAGCGTCTGGTGTGTCGATCCGCGCAACTCCAAATCATC 564
Qy 142 TyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrp 161

Db	172	CA	TCCC	AAG	CAAA	AGCG	TAG	CAT	ATG	CA	CA	AG	A	TAT	TGG	CGG	CTG	TAC	CGC	CTC	231
		:::				:															
Qy	45	Asp	Ala	Ala	Asn	Gly	A	g	T	r	P	ile	Pro	Leu	Leu	Asp	T	r	P	Val	60
		---				:															---
Db	232	AAC	---	GCC	AG	CA	G	ACT	CAT	G	AC	CG	CG	T	CA	CG	AT	TG	CG	AT	288
		---				:															---
Qy	61	Asn	Asn	Trp	Gly	Tyr	Asn	Gly	Val	Val	Val	Ser	Ile	Ala	Ala	Asp	Pro	Ile	Asn	Thr	Asn
		---				:									:						---
Db	289	CAC	AAC	TGG	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
		---				:									:						---
Qy	81	Val	Trp	Ala	Ala	Val	Gly	Met	Tyr	Thr	Asn	Ser	Trp	Asp	Pro	Asn	Asp	Gly	Ala	Ile	Leu
		---				:									:						---
Db	340	GTG	TAT	CCG	CA	G	T	CGG	CAT	GT	AT	AC	GA	CA	AG	T	CGG	AT	CCG	AG	T
		---				:									:						---
		---				:									:						---

[illegible][illegible]

Qy	Db
181 ThrGlyTyrGlnSerAspIleGlnGlyValValTpvValAlaPheAspLysSerSer	640 AACGGCTACCAACCGGACAAGCAGGCCCTCATGTGGGTTACGTTCACTACCCAGCATC

Qy 201 SerLeuGly-GlnAlaSerLysThrIlePheValGlyValAlaAsp 215
::: |||||
Db 700 ACGACCGGTGGCAGCCACGCTCGTATTTTGTCTGCCACGACTGAC 745

RESULT 4
CB905388 814 bp mRNA linear EST 02-JUL-2000
LOCUS CB905388
DEFINITION trico074xd02 T.reesei mycelial culture, Version 3 april Hypocrea

jecorina cDNA clone tric074xd02, mRNA sequence.
 CB905388
 ACCESSION

VERSION	CB905388.1	GI:30120046
KEYWORDS	EST.	
SOURCE	Hypocrea jecorina	(anamorph: Trichoderma reesei)
ORGANISM	Hypocrea jecorina	
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	
	Hypocreomycetidae; Hypocreales; Hypocreales; Hypocrea.	
REFERENCE	1 (bases 1 to 814)	
AUTHORS	Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,	
	Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,	
	Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchison, C.,	
	Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.	

TITLE Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus *Trichoderma reesei*
J. Biol. Chem. 278 (34), 31988-31997 (2003)
JOURNAL
MEDLINE 22803314
COMMENT 12788320
Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635

Email: Pforeman@genencor.com

```

FEATURES
source
Location/Qualifiers
1. .814
/organism="Hypocrea jecorina"

```

```

/mo1_cfp=nan.
/strain="QM6a"

```

```

FEATURES
source
1. .814
Location/Qualifiers
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"

```



```
/db xref="taxon:51453"
/clone="tric074xd02"
/dev stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/notes=vector: pREP3y; Site_1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Alignment Scores:
Pred. No.: 3.21e-43 Length: 814
Score: 587.50 Matches: 112
Percent Similarity: 68.52% Conservative: 36
Best Local Similarity: 51.85% Mismatches: 58
Query Match: 14.56% Indels: 10
Db: 6 Gaps: 4

US-09-917-376-3 (1-740) x CB950388 (1-814)
Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyValPhe 24
Db 179 TTTTCATGGAGAACGTCUACGTCGGCGCGCGCGGCTTCGTCGCCGGATCATCTTC 238
Qy 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44
Db 239 CATCCCAAGCAAAAGCGGTAGCATATGCAGACAGATATTGGCGGCTGTACCGCCTC 298
Qy 45 AspAlaIleAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60
Db 299 AAC---GCCGACGACTCATGACCGCGCTCAGGATGGATGTGCTGAATGCGGGCTGG 355
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 356 CACAACTGG-----GGCATGACGCTGTTGGCTTGTATCCGCGAGCATCAAAAG 406
Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
Db 407 GTGTATGCCGCGAGTCGGATGATACGACAGCTGGGATCCGAGTAATGGAGCCATCAT 466
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120
Db 467 CGCTCGTCAGACCGCGCGCAACGCTGTCTTCCAACTTGGCCTTCACAGTCGGGGGT 526
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAspAsnIle 140
Db 527 AACATGTCAGACCGCGAGCCGAGAGCGTCTGGCTGTGATCCGCGCCACCTCCAAACATC 586
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160
Db 587 ATCTACTTGTGTCTCGCTCAGACACGCGCTCTGGAGTCTACCGACGGCGGTGACC 646
Qy 161 TrpSerGlnMetThrAsnAspProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 647 TTTTCCAAGGCTCGTCGCTCAGCGCCCTCGGAGCTCCCTCCAGACCCGAGTGATTCC 706
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200
Db 707 AACGGCTACCAACCGCGACAAAGCGCTCATGTGGGTGTACGTTCGACTACCCAGCATC 766
Qy 201 SerLeuGly-GlnAlaSerLysThrIlePheValGlyValAlaAsp 215
Db 767 ACGACCGGTGGACGCCAGCTCGTAATTTTGTGTCGCCAGTGC 812

RESULT 5
CF880713 LOCUS
DEFINITION CF880713.1 bp mRNA linear EST 31-OCT-2003
Hypocrea jecorina cDNA clone tric082xn19, mRNA sequence.
ACCESSION CF880713
VERSION CF880713.1 GI:38135395
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
1 (bases 1 to 707)
Diener, S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,
Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,
Ward, M. and Dean, R.A.
Characterization of the protein processing and secretion pathways
in a comprehensive set of expressed sequence tags from Trichoderma
reesei
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph_dean@ncsu.edu
Seq primer: LT-F1 primer.
Location/Qualifiers
source
1..707
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric082xn19"
/dev stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/notes=vector: pREP3y; Site_1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Alignment Scores:
Pred. No.: 1.58e-27 Length: 707
Score: 415.50 Matches: 92
Percent Similarity: 55.51% Conservative: 34
Best Local Similarity: 40.53% Mismatches: 90
Query Match: 10.23% Indels: 11
Db: 7 Gaps: 3

US-09-917-376-3 (1-740) x CF880713 (1-707)
Qy 377 LysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 396
Db 33 CGCTCGCTGGATGATTGAGTCTCTCAGATTGACCCACGACAGCAACCACTGGCTC 92
Qy 397 TyrGlyThrGlyValaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGly 416
Db 93 TAGCGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACCACCTGGGACACGCGCCAC 152
Qy 417 GlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIle 436
Db 153 AATGTGTCAATCCAATCACTGGCAGACGCGCATCGAGGATTTCTCGTCCAGGACCTGGCC 212
Qy 437 SerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 456
Db 213 TCTGCACCGCGGGAAGCGAGCTATTGGCCGAGCTCGGAGCTCGGACGACACACCGCTTACCTTT 272
Qy 457 Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr 473
Db 273 GCCAGCAAGAACGACCTCGGACATCGCGCAGACGGTCTGGGCAACGCCCATCGGCC 332
Qy 474 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer 493
Db 333 ACCTCGACGAGCGTCCGATACGCCGGGAACCTCGTCAGAGCGTCTCGCGCGTCGCGCAC 392
Qy 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 513
Db 393 ACCGCGCGCACCGCAACAG-----GTGGCCATCTCGTCCGCGCGCGCGCG 437
Qy 514 AsnTrpPheGlnGlySerGluProGlyGlyValThrGlyValThrGlyGlyThrValAlaIle 533
```


Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph_dean@ncsu.edu


```

FEATURES
  source
    Seq primer: LT-F1 primer.
    Location/Qualifiers
      1..693
        /organism="Hypocrea jecorina"
        /mol_type="mRNA"
        /strain="QM6a"
        /db_xref="taxon:51453"
        /clone="trio029xol3"
        /dev_stage="mycelia"
        /clone_lib="T.reesei mycelial culture, Version 6 October 2003"
        /note="Vector: pREP3Y, Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."
ORIGIN
  source
    1..751
      /organism="Hypocrea jecorina"
      /mol_type="mRNA"
      /strain="QM6a"
      /db_xref="taxon:51453"
      /clone="trio016xm14"
      /dev_stage="mycelia"
      /clone_lib="T.reesei mycelial culture, Version 6 October 2003"
      /note="Vector: pREP3Y, Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."
ORIGIN
  Alignment Scores:
    Pred. No.: 1,57e-22 Length: 751
    Score: 361.50 Matches: 88
    Percent Similarity: 51.61% Conservative: 40
    Best Local Similarity: 35.48% Mismatches: 103
    Query Match: 8.96% Indels: 17
    DB: 7 Gaps: 6
  US-09-917-376-3 (1-740) x CF868882 (1-751)
  QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe 24
  Db 98 TTTTCATGGAAGACGTCAAGCTCGCGCGCGCGCTTCGTCCCGGCATCATCTTC 157
  QY 25 AsnGluGlyAlaProGlyLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44
  Db 158 CATCCAGACAAAGAGGTAGCATATGCACGACAGATATTGGCGGCTGTACCGCCTC 217
  QY 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60
  Db 218 AAC---GCCGACGACTCATGTGACCGCGCTCAGGATGGATGTCTGATATGTCGCGCTGG 274
  QY 61 AsnAsnTrpGlyTyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnLys 80
  Db 275 CACAACTGG-----GGCATCGACCTGTGTGGCTTGATCCGCATGACGATCAAAAG 325
  QY 81 ValTrpAlaAlaValAGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
  Db 326 GTGTATGCGCGACGCGCGATGATACGACACCTGAGATCTTACTATTGGACCCATCAIT 385
  QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120
  Db 386 CCTCGTCAGACACGCGGCAAGTGTCTTCCAACTTGCCCTTCACAGTCAGGGGT 445
  QY 121 AsnMetProGlyArgGlyMetGlyuArgLeuAlaValAspProAsnAsnAspAsnIle 140
  Db 446 AACATGCCAGACTCGGAGCGCGAGAGGCTGTGGCTGTCAATCCCTCCAACTACACATC 505
  QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeu---TrpArgSerThr 155
  Db 506 ATCTACTGTGGCGCTCGCTCAAGAAACAGCCTCTCTGGAAGTCTACG 551
  RESULT 8
  CF868882
  LOCUS
  DEFINITION
    trico16xm14.b1 T.reesei mycelial culture, Version 6 October 2003
  Hypocrea jecorina cDNA clone trico16xm14, mRNA sequence.
  ACCESSION
    CF868882
  VERSION
    CF868882.1 GI:38123564
  KEYWORDS
    EST.
  SOURCE
    Hypocrea jecorina (anamorph: Trichoderma reesei)
  ORGANISM
    Hypocrea jecorina
    Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
    Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
  REFERENCE
    1 (bases 1 to 751)
    Diener,S.E., Dunn-Coleman,N., Foreman,P., Houfek,T.D.,
    Teunissen,P.J.N., van Solingen,P., Dankmeyer,L., Mitchell,T.X.,
  TITLE
    JOURNAL
    COMMENT
  FEATURES
    source
    1..751
      /organism="Hypocrea jecorina"
      /mol_type="mRNA"
      /strain="QM6a"
      /db_xref="taxon:51453"
      /clone="trico16xm14"
      /dev_stage="mycelia"
      /clone_lib="T.reesei mycelial culture, Version 6 October 2003"
      /note="Vector: pREP3Y, Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."
ORIGIN
  Alignment Scores:
    Pred. No.: 1,57e-22 Length: 751
    Score: 361.50 Matches: 88
    Percent Similarity: 51.61% Conservative: 40
    Best Local Similarity: 35.48% Mismatches: 103
    Query Match: 8.96% Indels: 17
    DB: 7 Gaps: 6
  US-09-917-376-3 (1-740) x CF868882 (1-751)
  QY 407 AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGly 426
  Db 15 CACGATTTCAACCAATGGGACACGCGCCCAATGTGTCAATCAATCACTCTGGCAGACGGC 74
  QY 427 LeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSer 446
  Db 75 ATCGAGGAATTCCTCGTCAGGACCTGGCTCTGCACCCGCGGAGAGCGAGTATTGGCC 134
  QY 447 AlaLeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValPro 463
  Db 135 GCAGTCGGAGACGACACACGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCG 194
  QY 464 SerThrIlePheThrSerProValPheThrGlyThrSerValAspTyrAlaGluLeu 483
  Db 195 CAGACGCTGTGGGAACGCGCCACATGGGCCACCTCGACGAGCGTCCACTACGCGCGGAAC 254
  QY 484 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503
  Db 255 TCGGTCAAGAGCGTGTCTCGCTCGGCACACCGCGCGGACGACAG-----302
  QY 504 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523
  Db 303 ---GTGGCCATCTCGTCCGACGCGCGCGCGCATCGTGGAGCATCGACTACGCGCGCACACG 359
  QY 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543
  Db 360 TCCATGAACGCGGCGACGCGTGGCTATTTCGCGCGACGCGGACACGATCTCTCTGTCTGAC 419
  QY 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563
  Db 420 GCCTCGTCGCGC-----GTGCGAGCGCTCGCAGTTCAGGGGAGCTTGTGCTCGCTC 470
  QY 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583
  Db 471 TCGAGCTTCCCGCGCGGCGCGCTCATCGCTCGGACACAGAGACCAACACGCGTCTTCTAC 530
```


QY 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603
DB 531 GCCGCTCCGATCGACCTTTTACGTACAGGACACCGGACGAGCTTC-----581
QY 604 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 621
DB 582 ACGCCGGG---CCCAAGCTGGCAGCGCAGGACGATCCGGATATCGTCTCACCCG 638
QY 622 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 641
DB 639 ACCACCGCGGCGACGTGTATGTCACCGACGTCGCGATATCGCTCCACAGACTCG 698
QY 642 GlySerSerTrpSerAlaIleThr 649
DB 699 GGCACGACCTTTGGCCAAAGTCTCC 722

RESULT 9
CB898982
LOCUS
DEFINITION
trio016xm14 T. reesei mycelial culture, Version 3 april Hypocrea

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM
Hypocrea jecorina (anamorph: Trichoderma reesei)

REFERENCE
AUTHORS
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,

TITLE
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei

J. Biol. Chem. 278 (34), 31988-31997 (2003)
22803314
12788920

COMMENT
Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA

TEL: (650) 846-7635
FAX: (650) 621-7817
Email: pforeman@genencor.com
Seq primer: Lt-F1 primer.

FEATURES
source

1..803
/organism="Hypocrea jecorina"
/mol_type="mRNA"

/strain="QM6a"
/db_xref="taxon:51453"
/clone="trio016xm14"
/dev_stage="mycelia"
/clone_lib="T. reesei mycelial culture, Version 3 april"
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:

Pred. No.: 1,74e-22 Length: 803
Score: 361.50 Matches: 88
Percent Similarity: 51.61% Conservative: 40
Best Local Similarity: 35.48% Mismatches: 103
Query Match: 8.96% Indels: 17
DB: 6 Gaps: 6

US-09-917-376-3 (1-740) x CB898982 (1-803)

QY 407 AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGly 426

DB 67 CACGATTTCCACCACTGGGACACCGCCCAATGTGTCAATCCATCATCTGGCAGCGGC 126

QY 427 LeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSer 446
DB 127 ATCCAGGAATTCCTCGTCAGGACCTTGGCTCTGCACCCGCGGAAAGAGCTATTGGCC 186
QY 447 AlaLeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValPro 463
DB 187 GCAGTCGGAGACACACCGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCG 246
QY 464 SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 483
DB 247 CAGACGCTCTGGCAACGCCACATGGCCACCTCGACGAGCGTCGACTACGCGGGAAC 306
QY 484 AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 503
DB 307 TCGTCAAGAGCGTCGTCGCGTGGCAACACCGCGCGCACGACAG-----354
QY 504 HisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGluProGlyGly 523
DB 355 ---GTGGCCATCTCGTCCGACGCGCGGCGACGTGGAGCATCGACTACGCGCGCACAG 411
QY 524 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTtpAlaPro 543
DB 412 TCCATGACCGCGCGCACGCTGTCCTATTTCGCGCGCGCGCACGATCTCTGTGTCGACC 471
QY 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer 563
DB 472 GCTCTGTCGCGC-----GTGACGCTCGCAGTTCAGGCGCAGCTTTCCTCCGTC 522
QY 564 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 583
DB 523 TCGAGCTCGCCGCGCGCGCGTCATCGCTCGACAGAAAGAACCAACAGCGCTTCTTAC 582
QY 584 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 603
DB 583 GCGCGCTCGGATCGACCTTTTACGTGACGACGACACCGCGCAGCCTTC-----633
QY 604 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 621
DB 634 ACGCCGCGG---CCCAAGCTGGCAGCGCAGGACGATCCGGATATCGCTGCTCACCCG 690
QY 622 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 641
DB 691 ACCACCGCGGCGACGTGTATGTCGACCGACGTCGCGATATTCGCTCCACAGACTCG 750
QY 642 GlySerSerTrpSerAlaIleThr 649
DB 751 GGCACGACCTTTGGCCAAAGTCTCC 774

RESULT 10

CF866388

LOCUS

DEFINITION

trio006xe08.b1 T. reesei mycelial culture, Version 6 October 2003

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CF866388 738 bp mRNA linear EST 31-OCT-2003
trio006xe08.b1 T. reesei mycelial culture, Version 6 October 2003
Hypocrea jecorina cDNA clone trio006xe08, mRNA sequence.

CF866388

CF866388.1 GI:38121014

EST.

Hypocrea jecorina (anamorph: Trichoderma reesei)

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

1 (bases 1 to 738)

Diener, S.E., Dunn-Coleman, N., Foreman, P., Houfek, T.D.,

Teunissen, P.J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K.,

Ward, M. and Dean, R.A.

Characterization of the protein processing and secretion pathways

in a comprehensive set of expressed sequence tags from Trichoderma

reesei

FEMS Microbiol. Lett. 230 (2), 275-282 (2004)

Contact: Ralph A. Dean

Fungal Genomics Laboratory

North Carolina State University

Campus Box 7251, Raleigh, NC 27695, USA

Tel: 919-513-0020
 Fax: 919-513-0024
 Email: ralph.dean@ncsu.edu
 Seq primer: LT-F1 primer

FEATURES

source
 1..738
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"
 /db_xref="taxon:51453"
 /clone="tric006xe08"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 6 October 2003"
 /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:
 Pred. No.: 2,19e-20 Length: 738
 Score: 338.00 Matches: 89
 Percent Similarity: 51.97% Conservative: 43
 Best Local Similarity: 35.04% Mismatches: 102
 Query Match: 8.37% Indels: 20
 DB: 7 Gaps: 9

US-09-917-376-3 (1-740) x CF866388 (1-738)

QY 422 ProMetValLys--GlyLeuGluThrAlaValAspLeuSerProProSer 440
 Db 2 CCACGGTCCGAGCGCATCGAGAAATTCCTCGTCCAGGACCTGGCTCTGCACCGGC 61
 QY 441 GlyAlaProLeuSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla----- 457
 Db 62 GGAAGCGAGCTATTGCGCGCAGTCGGAGACGACAAACGGCTTCACCTTTGCCAGCAAAAC 121
 QY 458 AspValThrAlaValProSerThrPheThrSerProValPheThrGlyThrSer 477
 Db 122 GACCTCGGAGATCCCGCGAGACGGTCTGGGACGCCACATGGCGCACCTCGACGAGC 181
 QY 478 ValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSer 497
 Db 182 GTCGACTACGCGGGAACCTCGTCAAGACGCTCGTCCGGTCCGCAACACCCCGCGCAG 241
 QY 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517
 Db 242 CAACAG-----GTGGCCATCTCGTCGACGGCGCGCGCAGCTGGAGCATC 286
 QY 518 GlySerGluProGlyGlyValThrGlyThrValAlaAlaSerAlaAspGlySer 537
 Db 287 GACTACGGCGCGCAGACGCTCCATGAAACGGCGGACGGTGGCTATTCCGGCCGACGGCGAC 346
 QY 538 ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557
 Db 347 AGCATCTCTGTCTGACCGCCCTCGTCCGC-----GTGACGGCTCGGAGTTCAG 397
 QY 558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577
 Db 398 GGCAGCTTGCTCTCGTCTCGAGCCTCGCGCGCGCGCGCTCATCGCTCGGACCAAGAAG 457
 QY 578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597
 Db 458 ACCAACAGCGCTTCTTACGCGCGCTCCGGATCGACCTTTTACGTGACGAAGACACCGGC 517
 QY 598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616
 Db 518 AGCAGCTTC-----ACGCGCGGG-----CCCAAGCTGGCGAGCGCGAGCATCCGG 565
 QY 617 ---PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu 635
 Db 566 GATATCGCTGCTCACCGACACCCCGGGGCGAGCTTGATGTCTCGACCGACCGCTCGGCATA 625

QY 636 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla 654
 Db 626 TTCCGCTCCACAGCTCGCGCAGGACCTTTGGCCAAGTCTCCACCGCCTGCCAACACC 685
 QY 655 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668
 Db 686 TACCAGATCGCCCTGGGTGGGCTCA---GGCTCGAAGTGG 724

RESULT 11

CB896354 794 bp mRNA linear EST 02-JUL-2003
 LOCUS CB896354
 DEFINITION jecorina cDNA clone tric006xe08, mRNA sequence.

ACCESSION CB896354

VERSION 1

KEYWORDS EST

SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)

ORGANISM Hypocrea jecorina

REFERENCE 1 (bases 1 to 794)
 AUTHORS Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
 Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
 Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
 Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.

TITLE Transcriptional regulation of biomass-degrading enzymes in the
 filamentous fungus Trichoderma reesei

JOURNAL J. Biol. Chem. 278 (34), 31988-31997 (2003)

MEDLINE 22803314

PUBMED 12788920

COMMENT

Contact: Pamela K. Foreman

925 Page Mill Road, Palo Alto, CA 94304, USA

Tel: (650) 846-7635

Fax: (650) 621-7817

Email: Pforeman@genencor.com

Seq primer: LT-F1 primer.

FEATURES

source

1..794

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/strain="QM6a"

/db_xref="taxon:51453"

/clone="tric006xe08"

/dev_stage="mycelia"

/clone_lib="T.reesei mycelial culture, Version 3 april"

/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial

culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:

Pred. No.: 2,45e-20 Length: 794

Score: 338.00 Matches: 89

Percent Similarity: 51.97% Conservative: 43

Best Local Similarity: 35.04% Mismatches: 102

Query Match: 8.37% Indels: 20

DB: 6 Gaps: 9

US-09-917-376-3 (1-740) x CB896354 (1-794)

QY 422 ProMetValLys--GlyLeuGluThrAlaValAsnAspLeuSerProProSer 440

Db 58 CCACGGTCCGAGCGCATCGAGAAATTCCTCGTCCAGGACCTGGCTCTGCACCGGC 117

QY 441 GlyAlaProLeuSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla----- 457

Db 118 GGAAGGAGCTATTGCGCGCAGTCGGAGACGACAAACGGCTTCACCTTTGCCAGCAAAAC 177

QY 458 AspValThrAlaValProSerThrPheThrSerProValPheThrGlyThrSer 477

Db 178 GACCTCGGAGATCGCGCGAGCGGTCTGGGCAACGCCACATCGGGCCACCTCGACGAGC 237

478 ValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSer 497
|||||
Db 238 GTGACTACCGCGGAACCTCGTCAAGAGCTGTCGCGTCCGCAACACCGCGGACG 297
|||||
Qy 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517
|||||
Db 298 CAACAG-----GTGCGCATCTCGTCCGACGCGCGCGCGACGTCGGAGCATC 342
|||||
Qy 518 GlySerGluProGlyGlyValThrGlyThrValAlaAlaSerAlaAspGlySer 537
|||||
Db 343 GACTACGCGCGCACACGCTCATGAACGCGCGCAGCGTGGCTATTCGCGCGACGCGAC 402
|||||
Qy 538 ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557
|||||
Db 403 ACGATCTCTGTCGACCGCTCGTCGCGC-----GTGACGCGTCGCAGTTCAG 453
|||||
Qy 558 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 577
|||||
Db 454 GCGAGCTTTCCTCGTCTCGAGCTCGCGCGCGCGCGTTCATCGCTCGACAGAAG 513
|||||
Qy 578 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 597
|||||
Db 514 ACCAACACGCTTCTACGCGCGCTCGGATCGACCTTTTACGTCAGCAAGACACCGCG 573
|||||
Qy 598 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet--- 616
|||||
Db 574 AGCAGCTTC-----ACGCGCGG---CCCAAGCTGGCGCGCGGACGAGCATCGG 621
|||||
Qy 617 ---PheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu 635
|||||
Db 622 GATATCGTGTCTACCCACCGACCGCGCGCGCGTGTATGTCGACCGCGTCCGCATA 681
|||||
Qy 636 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla 654
|||||
Db 682 TTCCGCTCCACAGATCGCGCGACGCTTTGGCCAAAGTCTCCACCGCGCTGACCAACACC 741
|||||
Qy 655 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668
|||||
Db 742 TACCAGATCGCCCTCGGTGTGGGTCA---GGCTCGAACTGG 780
|||||

BF072664 546 bp mRNA linear EST 18-OCT-2000
NC3MH7T3 Subtracted Mycelial Neurospora crassa cDNA clone SM3H7 5'
similar to avicelase III, Aspergillus aculeatus, mRNA sequence.

BF072664
VERSION BF072664.1 GI:10866169
KEYWORDS EST.
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
REFERENCE Neurospora crassa
AUTHORS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
1 (bases 1 to 546)
Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,
Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Bluey, S.,
Cushing, T., Ertter, A., Fleharty, M., Gorman, M., Judson, K.,
Miller, R., Ortega, J., Pavlova, I., Perea, J., Todisco, S.,
Trujillo, R., Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S.
and Natvig, D.O.

Expressed sequences from conidial, mycelial, and sexual stages of
Neurospora crassa
Fungal Genet. Biol. 21, 348-363 (1997)
97435549
9290248
Contact: Natvig, D.O./Nelson, M.A.
Department of Biology
University of New Mexico
Castetter Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304
Email: ngp@biology.unm.edu.
Location/Qualifiers
1. .546

FEATURES
source

/organism="Neurospora crassa"
/mol_type="mRNA"
/strain="74-OR23-IV A (FGSC 2489)"
/db_xref="taxon:5141"
/clone="SM3H7"
/sex="Mating type A"
/tissue_type="Mycelium"
/dev_stage="Mycelium"
/lab_host="E. coli"
/clone_lib="Subtracted Mycelial"
/note="Vector: pBluescript SK (-); Site 1: EcoRI; Site 2:
XhoI; 2% sucrose for 24 hours. cDNA directionally cloned
into pBluescript SK(-) using the Uni-ZAP XR vector system
(Stratagene, La Jolla, CA). previously identified highly
expressed clones were subtracted from this library."

ORIGIN

Alignment Scores:
Pred. No.: 1.98e-18 Length: 546
Score: 314.50 Matches: 72
Percent Similarity: 54.26% Conservative: 30
Best Local Similarity: 38.30% Mismatches: 75
Query Match: 7.79% Indels: 11
DB: 2 Gaps: 3

US-09-917-376-3 (1-740) x BF072664 (1-546)

Qy 358 SerAlaGluProTrpLeuThrPheGlyValGlnProAsnProValProSerProLys 377

Db 5 TCGGCACGAGCTCGATCGAGACTGGCTTCCTTTTCAG-----GATACCAAGCAC 55

Qy 378 LeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr 397

Db 56 CTCGCTGATGATCATCAATCCCTCGAGATCAACCCCTCGACAGCACCATTTGGCTCTAC 115

Qy 398 GlyThrGlyAlaLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGln 417

Db 116 GCCACGGTCTCACTGTTTACGGCGGCACGACCTGACCAAGTGGGACACCGTCCACAAC 175

Qy 418 IleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuSer 437

Db 176 GTGACATCCAGTCTTGGCGCTTGGCATCGAAGAAATGGCTGTCTCGCTCGCTCC 235

Qy 438 ProProSerGlyAlaProLeuLeuSerAlaLeuGlyAspLeuGlyGlyPheThr----- 455

Db 236 GCCCGCGTCTCGAGCTTCTCGCGCGCTCGGTGACGACCTGCGGCTTCACCTTCAAG 295

Qy 456 ---HisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 474

Db 296 TCCAGCTCGACCTCGGACCTCTCCCAAGACACCTCGGATGACGCCCCAGTGGGCCAGC 355

Qy 475 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPhe 494

Db 356 TCCGCGGACCTCGACTACGCGCGCAACAAAGCCGCAACGCTGTGCGCATCGGCTCGGC 415

Qy 495 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyLysAsn 514

Db 416 TCCGCGCGCAGCAG-----GTGCGCGTATCTCTCGGACGCGCGGCGCTCG 460

Qy 515 TrpPheGlnGlySerGluProGlyGlyValThrGlyGlyThrValAlaAlaSerAla 534

Db 461 TGGCATGTCTACCAACGCGCAGCACCAACAGCAGCGGCTCGCTCTCTCTGCC 520

Qy 535 AspGlySerArgPheValTrpAla 542

Db 521 GAGCGCGACCATCGTCTGGTCT 544

RESULT 13

BU639045

LOCUS

DEFINITION

mgcw011xD18f.b RCW Lambda Zap Express Library Magnaporthe grisea

CDNA clone mgcw011xD18 5', mRNA sequence.

BU639045

ACCESSION

BU639045

389 bp mRNA linear EST 06-MAY-2003


```

VERSION      BU639045.2  GI:30397224
KEYWORDS     EST
SOURCE       Magnaporthe grisea (anamorph: Pyricularia grisea)
ORGANISM     Magnaporthe grisea
REFERENCE    1  Ebbole,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P., Lu,G.,
AUTHORS      Bhatteirai,K. and Dean,R.A.
TITLE        Expressed sequence tags from the rice blast fungus, Magnaporthe
              grisea
JOURNAL      Unpublished (2002)
COMMENT      On Sep 30, 2002 this sequence version replaced gi:23351371.
              Contact: Ebbole DJ
              Department of Plant Pathology & Microbiology
              Texas A&M University
              Peterson Bldg, MS2132, College Station, TX 77843-2132, USA
              Tel: 979 845 4831
              Fax: 979 845 6483
              Email: d-ebbole@tamu.edu
              Chromatogram file of this sequence is available, see contact
              person; Best nr hit (April. 22, 2003) gb|EAA29333.1| hypothetical
              protein [Neurospora crassa] 206 4e-53
              PCR Primers
              FORWARD: T3 primer
              BACKWARD: T7 primer
              Plate: mgc011 row: D column: 18
              Seq primer: T3.

FEATURES     source
             1..389
                Location/Qualifiers
                /organism="Magnaporthe grisea"
                /mol_type="mRNA"
                /strain="CP987"
                /db_xref="taxon:148305"
                /clone="mgc011xD18"
                /sex="Mat1-1 hermaphrodite"
                /tissue_type="Mycelium"
                /dev_stage="Day 5 post-inoculation"
                /clone_lib="RCW Lambda Zap Express Library"
                /note="Vector: phluescript excised from Lambda Zap
              Express; Site 1: EcoRI; Site 2: XhoI; Day 5
              post-inoculation mRNAs prepared from Magnaporthe grisea
              grown at 23C in the dark with constant gyratory shaking
              100 rpm in Vogel's minimal medium containing 0.5% isolated
              rice cell walls as the sole carbon source. Library
              provided by Sheng-Cheng Wu. Sequences were processed by
              one of two methods. Where a full-length alignment to the
              M. grisea genome sequence was available, the EST sequence
              was trimmed according to the alignment, otherwise sequence
              quality was assessed using phredphrap version 991019 and
              trimmed according to phd files (0.05) and for vector
              seqs."

ORIGIN
Alignment Scores:
Pred. No.:      6.41e-18      Length:      389
Score:          306.50      Matches:      55
Percent Similarity: 66.67%      Conservative: 27
Best Local Similarity: 44.72%      Mismatches: 38
Query Match:    7.59%      Indels:      3
DB:             5          Gaps:      1

US-09-917-376-3 (1-740) x BU639045 (1-389)

Qy 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353
Db 5  GGGGCTACATGTCAGATCTTGGAGTGGACAGTATCCCGATCGACATGACGACTACTAC 64
Qy 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProVal 373
Db 65 TCGATCAACACAGACAGCGCCCTGGATAGTCTGGC-----CATCTTCGAGG 115
Qy 374 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393

```

```

Db 116 GATTCGAAGCGCCTTGGTTGGATGATCGAGGCCCTTGAATCGACCCCTGCGATCGTAC 175
Qy 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIlystTrpAsp 413
Db 176 CATTTGGCTTTACGGAACTGGCCTCTCGATCTTTGGGGGACATGACTTGGACCAATCGGAC 235
Qy 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsn 433
Db 236 ACGGTTCAATACATATCGATTCACTCTTGGCCGATGTGTGCGAAGAACTGCAGTTCTG 295
Qy 434 AspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly 453
Db 296 GAGATGGCTTCAGCGCCCGAGGCTCCGAACTACTTGGCGCAATGGAGACATTACTGGA 355
Qy 454 PheThrHis 456
Db 356 TTCACATAC 364

RESULT 14
CF875815 713 bp mRNA linear EST 31-OCT-2003
LOCUS     trico39xxk06.b1 T.reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trico39xxk06, mRNA sequence.
ACCESSION CF875815
VERSION    CF875815.1 GI:38130497
KEYWORDS   EST.
SOURCE     Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM   Hypocrea jecorina
REFERENCE   1 (bases 1 to 713)
AUTHORS     Diener,S.E., Dunn-Coleman,N., Foreman,P., Houfek,T.D.,
              Teunissen,P.J.M., van Solingen,P., Dankmeyer,L., Mitchell,T.K.,
              Ward,M. and Dean,R.A.
TITLE       Characterization of the protein processing and secretion pathways
              in a comprehensive set of expressed sequence tags from Trichoderma
              reesei
JOURNAL     FEMS Microbiol. Lett. 230 (2), 275-282 (2004)
COMMENT     Contact: Ralph A. Dean
              Fungal Genomics Laboratory
              North Carolina State University
              Campus Box 7251, Raleigh, NC 27695, USA
              Tel: 919-513-0020
              Fax: 919-513-0024
              Email: ralph.dean@ncsu.edu
              Seq primer: LT-F1 primer.

FEATURES     source
             1..713
                Location/Qualifiers
                /organism="Hypocrea jecorina"
                /mol_type="mRNA"
                /strain="QM6a"
                /db_xref="taxon:51453"
                /clone="trico39xxk06"
                /dev_stage="mycelia"
                /clone_lib="T.reesei mycelial culture, Version 6 October
              2003"
                /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
              culture grown from 24 hrs to 6 days with varying Carbon
              and Nitrogen sources and concentrations."

ORIGIN
Alignment Scores:
Pred. No.:      1.11e-15      Length:      713
Score:          286.50      Matches:      81
Percent Similarity: 48.79%      Conservative: 40
Best Local Similarity: 32.66%      Mismatches: 106
Query Match:    7.10%      Indels:      21
DB:             7          Gaps:      10

US-09-917-376-3 (1-740) x CF875815 (1-713)

Qy 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 477

```


Db	10	GACCTCGGTACATCGCCGCGAGTTGCTGGGCAACGCCACATGGCGCCACCTCGACGAGC	69
Qy	478	ValAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSer	497
Db	70	GTGCACTACGGGGAACCTCGCTCAAGAGCGTCGTCGGCTCGGCAACACCGCCGCGACG	129
Qy	498	SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln	517
Db	130	CAACAG-----GTGGCCATCTCGTCGACGGCGCGCGACGTGGAGCATC	174
Qy	518	GlySerGluProGlyGlyValThrThrValAlaAlaSerAlaAspGlySer	537
Db	175	GACTACGGCGCGACACGTCATGAACGGCGGCGCGCTATTTCGCGCGACGCGAC	234
Qy	538	ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly	557
Db	235	ACGATCTCTGTGTCACCGCTCGTCGCGC-----GTGACGCGCTCGCAGTTCCAG	285
Qy	558	AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal	577
Db	286	GGCAGCTTTGCTCGCTCGAGCTGCGCGCGCGCGCTCATCGCTCGACAGAG	345
Qy	578	AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly	597
Db	346	ACCAACACGGCTTCTACGCCGCTCGGATCGACTTTTACGTTCAGCAAGACACCGGC	405
Qy	598	ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet---	616
Db	406	AGCAGCTTC-----ACGCGCGG---CCCAAGCTGGCGACGCGAGCATCCGG	453
Qy	617	----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu	635
Db	454	GATATCGCTGCTCACCGCACACCGCGCGCGCTTGTATGTCTCGACCGACGTCGCATA	513
Qy	636	TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla	654
Db	514	TTCCGCTCCACAGATCGGGGCGACGCTTTGGCAAGTCTCCACCGCTGACCAACACC	573
Qy	655	ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal	674
Db	574	TACCAGATCGCCCTGGGTGGGCTCA---GGCTCGAAGTGG---AACCTGTATGCTTC	627
Qy	675	GlyThrIleGlyGlyValThrGlyAla-----TyrArgSerAspAspCysGlyThrThr	692
Db	628	GGCACC-----GGCCCGCTCAGGGGCTCGCTCTACGCCAGTGTGAGACAGCGCGCTCC	681
Qy	693	TrpValLeuIleAsnAspAspGln	700
Db	682	TGGACGGACATTTCAGGGTCCCCAG	705
RESULT 15			
CB904767			
LOCUS			
DEFINITION			
trio039xk06 T. reesei mycelial culture, Version 3 april Hypocrea			
Jecorina cDNA clone trio039xk06, mRNA sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Hypocrea jecorina (anamorph: Trichoderma reesei)			
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
Hypocremycetidae; Hypocreales; Hypocreaceae; Hypocrea.			
REFERENCE			
AUTHORS			
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,			
Dunn-Coleman, N.S., Goeddebuur, F., Houfek, T.D., England, G.J.,			
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,			
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.			
TITLE			
Transcriptional regulation of biomass-degrading enzymes in the			
filamentous fungus Trichoderma reesei			
J. Biol. Chem. 278 (34), 31988-31997 (2003)			
JOURNAL			
MEDLINE			
PUBMED			
22803314			
12788920			

COMMENT			
Contact: Pamela K. Foreman			
Genencor Intl.			
925 Page Mill Road, Palo Alto, CA 94304, USA			
Tel: (650) 846-7635			
Fax: (650) 621-7817			
Email: Pforeman@genencor.com			
Seq primer: Lr-F1 primer.			
Location/Qualifiers			
source			
1..782			
/organism="Hypocrea jecorina"			
/mol_type="mRNA"			
/strain="QM6a"			
/db_xref="taxon:51453"			
/clone="trio039xk06"			
/dev_stage="mycelia"			
/clone_lib="T. reesei mycelial culture, Version 3 april"			
/notes="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial			
culture grown from 24 hrs to 6 days with varying Carbon			
and Nitrogen sources and concentrations."			
ORIGIN			
Alignment Scores:			
Pred. No.: 1.28e-15 Length: 782			
Score: 286.50 Matches: 81			
Percent Similarity: 48.79% Conservative: 40			
Best Local Similarity: 32.66% Mismatches: 106			
Query Match: 7.10% Indels: 21			
DB: 6 Gaps: 10			
US-09-917-376-3 (1-740) x CB904767 (1-782)			
Qy	458	AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer	477
Db	79	GACCTCGGTACATCGCGCGAGTTGCTGGCAACGCCACATGGCCACTCGACGAGC	138
Qy	478	ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer	497
Db	139	GTGCACTACCGCGGAACTCGGTCAAGAGCGTCTCCGCGCGCAACACCGCGGACG	198
Qy	498	SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln	517
Db	199	CAACAG-----GTGGCCATCTCGTCCGACGCGCGCGCGCGGCGGAGCATC	243
Qy	518	GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer	537
Db	244	GACTACGGCGCGACAGCTCATGAACGGCGGCGCGGCTTATTCGCGCGACGCGCAG	303
Qy	538	ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly	557
Db	304	ACGATCTCTGTGTCGACCGCTCGTCCGGC-----GTGACGCGCTCGCAGTCCAG	354
Qy	558	AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal	577
Db	355	GGCAGCTTTGCTCCGTCCTCGAGCTGCGCGCGCGCGCGCTCATCGCTCGGACAAG	414
Qy	578	AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly	597
Db	415	ACCAACAGCGCTTCTTACGCGCGCTCGGATCGACCTTTTACGTCAAGACACCGCGC	474
Qy	598	ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet---	616
Db	475	AGCAGCTTC-----ACGCGCGG---CCCAAGCTGGCGGCGGCGGACGATCCGG	522
Qy	617	---PheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu	635
Db	523	GATATCGCTGCTCACCCGACACCGCGGCGCGCTTGTATGTCTCGACCGACGCGATA	582
Qy	636	TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAla	654
Db	583	TTCCGCTCCACAGACTCGGCGCACACCTTTTGGCCAAAGTCTCCACCGCTGACCAACC	642
Qy	655	ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal	674


```

Db      643 TACCAGATCGCCCTGGGTGTGGCTCA---GGCTCGAACTGG---AACCTGTATGCCTTC 696
Qy      675 GlyThrIleGlyGlyValThrGlyAla-----TyrArgSerAspAspCysGlyThrThr 692
Db      697 GGCACC-----GGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCTCC 750
Qy      693 TrpValLeuIleAsnAspAspGln 700
Db      751 TGGACGGACATTCAGGGCTCCAG 774

```

Search completed: October 5, 2005, 21:09:27
 Job time : 7126.69 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 5, 2005, 07:47:30 ; Search time 352.231 Seconds
(without alignments)
3437.649 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRGIYVDIGGAPSG 740

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09917376/runat_04102005_164328_26925/app_query.fasta_1.2524
-DB=Issued_Patents_NA -QPMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=5125 -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CNC 1.1 516 @runat_04102005_164328_26925 -NCFU=6 -ICFU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=10 -XGAPOP=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	227.5	5.6	3129	US-09-252-991A-13873	Sequence 13873, A
C 2	227.5	5.6	8211	US-09-252-991A-13656	Sequence 13656, A
C 3	203	5.0	4188	US-09-252-991A-13774	Sequence 13774, A
C 4	201.5	5.0	4131	US-09-252-991A-13773	Sequence 13773, A
C 5	192	4.8	4411529	US-09-103-840A-1	Sequence 1, Appli
C 6	184	4.6	10317	US-09-902-540-1027	Sequence 1027, Ap
C 7	184	4.6	10317	US-09-902-540-1027	Sequence 2, Appli
C 8	181.5	4.5	11679	US-09-328-352-1377	Sequence 1377, Ap
9	177	4.4	29927	US-09-949-016-11814	Sequence 11814, A
10	177	4.4	29927	US-09-949-016-11814	Sequence 17474, A
11	177	4.4	29927	US-09-949-016-17475	Sequence 17475, A
12	177	4.4	37802	US-09-949-016-12639	Sequence 12639, A

13	175.5	4.3	4287	4	US-09-902-540-2590	Sequence 2590, Ap
14	175.5	4.3	16584	4	US-09-902-540-1119	Sequence 1119, Ap
15	170.5	4.2	25165	3	US-09-453-702B-39	Sequence 39, Appli
16	170	4.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
17	169	4.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
18	168.5	4.2	20113	4	US-09-902-540-1173	Sequence 1173, Ap
19	167.5	4.2	4236	4	US-09-902-540-5367	Sequence 5367, Ap
C 20	167.5	4.2	34662	4	US-09-902-540-1261	Sequence 1261, Ap
C 21	164	4.1	2319	4	US-09-252-991A-13875	Sequence 13875, A
C 22	163	4.0	5741	1	US-07-706-699-4	Sequence 4, Appli
23	163	4.0	5741	1	US-07-998-931-4	Sequence 4, Appli
24	160.5	4.0	47981	4	US-09-679-279-1	Sequence 1, Appli
25	159.5	4.0	3180	4	US-09-248-796A-20	Sequence 20, Appli
26	159.5	4.0	4647	4	US-09-252-991A-5730	Sequence 5730, Ap
27	159.5	4.0	10419	4	US-09-408-020-3	Sequence 3, Appli
28	159.5	4.0	42432	4	US-09-408-020-2	Sequence 2, Appli
C 29	159	3.9	13805	4	US-09-902-540-1083	Sequence 1083, Ap
30	157.5	3.9	2067	4	US-09-489-039A-2642	Sequence 2642, Ap
31	157	3.9	3155	4	US-09-710-279-3881	Sequence 3881, Ap
C 32	156.5	3.9	1860	4	US-09-252-991A-9781	Sequence 9781, Ap
C 33	156.5	3.9	5132	4	US-09-902-540-597	Sequence 597, App
34	156	3.9	26930	4	US-09-902-540-1228	Sequence 1228, Ap
35	155	3.9	72704	4	US-09-902-540-1273	Sequence 1273, Ap
C 36	155.5	3.9	2733	4	US-09-710-279-3107	Sequence 3107, Ap
37	155	3.8	2235	1	US-08-418-782-1	Sequence 1, Appli
38	155	3.8	2235	1	US-08-228-662-1	Sequence 1, Appli
39	155	3.8	2235	2	US-08-852-219-1	Sequence 1, Appli
40	154	3.8	3900	4	US-09-023-655-1420	Sequence 1420, Ap
C 41	154	3.8	16013	4	US-09-949-016-12988	Sequence 12988, A
42	153	3.8	2625	6	5457037-4	Patent No. 5457037
43	153	3.8	2625	6	5457037-4	Patent No. 5457037
44	153	3.8	3050	4	US-09-710-279-3661	Sequence 3661, Ap
45	153	3.8	3336	6	5457037-1	Patent No. 5457037

ALIGNMENTS

RESULT 1
US-09-252-991A-13873/c
Sequence 13873, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 13873
LENGTH: 3129
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13873

Alignment Scores:	6.58e-11	Length:	3129
Pred. No.:	227.50	Matches:	218
Score:	31.30%	Conservative:	90
Best Local Similarity:	22.15%	Mismatches:	360
Query Match:	5.64%	Indels:	319
DB:	4	Gaps:	52

US-09-917-376-3 (1-740) x US-09-252-991A-13873 (1-3129)

QY 20 AspGlyIleValPheAsnGluGlyAla---ProGlyIleLeuTyrValArgThrAspIle 38
Db 2993 AACGGCGTCGTCATCAGCGGCGGAGCGCGGACCGTGCACCTCACCAGTGCC 2934

QY 39 GlyGlyMetTyrArgTrpAspAlaalaasnGlyArgTrpIleProLeuLeuAspTrpVal 58
Db 2933 GGC-----GGCAACCGGATAGGCGAGGTCAACCGCGAC 2901
QY 59 GlyTrpAsnAsnTrpGlyTyr-----AsnGly-----ValVal 69
Db 2900 GGCAGCGCAACTGGAGTTCACGCGCGGACGCGCGGCAACGAGCACGGTGTATCGTC 2841
QY 70 SerIleAlaAlaAspProIle---AsnThrAsnLysValTrpAla-----Ala 84
Db 2840 GCCAGCGCACCGGACCGGCAATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2781
QY 85 ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu----- 100
Db 2780 GTGGCGCGCGCGCGCGGTGATCGATCCGAGCAACGCGCACCATCAGCGGCGCACCGCG 2721
QY 101 -----ArgSer 102
Db 2720 GAGCGCGGCGCAAGTGTCTCTCACCGACGCGCAACCGGCAACCGGCGGCGGCAACCCACC 2661
QY 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGly----- 119
Db 2660 GCCAGCGGCGGCGCAACTGGACCTTACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2601
QY 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArg 130
Db 2600 GTCAACCGCGTGGCGGACCGCTTGGGCGCAATACCGGCGGCGGCGGCGGCGGCGGCGGCG 2541
QY 131 LeuAlaValAspProAsn-----AsnAspAsnIleLeuTyrPhe 143
Db 2540 GACCGCGTGGCGCGGCAACCGCTTGTGTCAATCCGAGCAACGCGCAACCTGCTCAACGGT 2481
QY 144 GlyAlaProSerGlyLysGlyLeu-----Trp 152
Db 2480 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2421
QY 153 ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe 167
Db 2420 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2364
QY 168 ProAspValGlyThrTyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAsp 186
Db 2363 -----GGCACCGTGTCAACGTCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2316
QY 187 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln----- 204
Db 2315 -----GCTCCCGCTACCGACGCGTGGATTCTCGTGGCGGCGGCGGCGGCGGCGGCGG 2262
QY 205 ---AlaSerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProValPhe----- 221
Db 2261 CCGAGCAACGGTTCGGTGTATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2202
QY 222 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAla 232
Db 2201 GATGCAACGCAACCGGATGGCGAGTCCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2142
QY 233 ValProGlyAlaPro----- 237
Db 2141 ACTCCAGCATCCCGCTGCGGATGGCGAGTGGTTCACGCGGCGGCGGCGGCGGCGGCGGCG 2082
QY 238 -----ThrGlyPheIleProHisLysGlyVal 246
Db 2081 AATGTCAACAGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2022
QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrThrGlyProTyr--- 265
Db 2021 ATCGATCCGACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1962
QY 266 -----AspGlySerSerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp 282
Db 1961 CTCACCGATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1902
QY 283 ThrArgIleSerProValPro-----SerThrAsp 292

Db 1901 ACGTTCACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1842
QY 293 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 310
Db 1841 CCGCGCGGCAATACACGCGGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1782
QY 311 -----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324
Db 1781 CCGGTGATCAATCCGAGCAACGGAGTCGTATCAGCGGTACGCGGGAACCGGCGGCGGCGG 1722
QY 325 ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr 338
Db 1721 GTGATCTCT-----ACCGAGCGCAACCGGCAACCGGATCGGCGGCGGCGGCGGCGG 1671
QY 339 ArgIleTrpAspTrpThr-SerTyrProAsnArg-----SerIle 351
Db 1670 GGCAGCGCAAGTGGGCTTTCACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1611
QY 351 uArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPr 371
Db 1610 CCGTGG-----CCGAGGCGCGCGGCGGCAACAGCAGTC-----CC 1572
QY 371 oProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAs 391
Db 1571 ACCAGCGCA---CGTCTGACTCGTGGCGGCGGCA--GCGCGGTGATCGATCCGAGCAA 1516
QY 391 nSerAspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAs 408
Db 1515 CGGTAGC---GTGATCGCGGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1464
QY 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGl 428
Db 1463 -----GACGGCAACGGCAAC-----CCGATCGGCGGCGGCGGCGGCGGCGG 1429
QY 428 uGluThrAlaValAsnAspLeuIleSerProSerProSerGlyAlaProLeu----- 444
Db 1428 CGATGGCAGCGCAACTGGAGCTTCACGCCC-----GGCAGCGCGCTGCTCAATGGCAC 1375
QY 445 -----IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAl 461
Db 1374 GGTGGTCAATCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1315
QY 461 aValProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal-- 478
Db 1314 GGTGACTCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1255
QY 479 -AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 498
Db 1254 CGGTACCGCGGAGCGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1195
QY 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 515
Db 1194 CCAGGCG-----ACCGCGGATGGCAGCGCACTGGAGCTTCAC 1156
QY 516 -----PheGlnGlySerGluProGlyGl 523
Db 1155 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1096
QY 523 yValThrThrGlyGlyThr----- 529
Db 1095 CAATACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1036
QY 530 -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAl 542
Db 1035 CAACCGGAGCAACGCGAGCGGTGATCGCGGTACCGGGAAGCGGCGGCGGCGGCGGCGGCG 976
QY 542 aProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTr 560
Db 975 CACCGAGCGGCGGCGGCAACCGGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 919
QY 560 pAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer----- 574


```

Db 918 GAGCTTCACGGCCGCGCAGCGCGTGGCAACAGCGCTCGGTGATCAATCGCTGGGCCAGGA 859
Qy 575 -----AspArgValAsnProLysThr 581
Db 858 CGCCGCGCGCAACACACAGCGCGCGCGCAGCACCACCGTGGACTCGGTAGCCCGCGCCAC 799
Qy 581 r---PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer----- 593
Db 798 CCGCGTGTCTGATCCGAGCAACCGTACGCTGATCAGCGGTACCGCGTACCGCGAAGCGCGGGCCAC 739
Qy 594 -----ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSers 610
Db 738 GGTGATCTTACCGAGCGCGC-----GGCAACCGATACCGCAGCGCCACCGCGATGG 685
Qy 610 erglyAlaValGlyValMetPheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuA 630
Db 684 CAGCGCAACTGGAGCTTCACTCCGGGCACA-CGGCTGACCAACGCGCAGCGTGTATCATG 626
Qy 630 laAlaSerSerGlyLeuTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaI 648
Db 625 CGGTGGCCAGGACCGCGCGCAACACCGCGGTCCCGTCAGCACCACACATGGAGACCGCG 566
Qy 648 leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSert 668
Db 565 TGGCCCCCGCACCCCGGTGTATCGACCGAGCATGGTGTCAAACTCAGCGGCACCGCGCG 506
Qy 668 yrProAlaValPheValValGlyThrIleGlyGly----- 679
Db 505 AACCGCGGTCCGGTGTATCTCCAGCATGGCAACCGCATCGCCAGCACCTCTCG 446
Qy 680 -----ValT 681
Db 445 CCGAGGTAGCGGCAACTGGACCTTACACCGCGGCACCGCGTGGCCAAACCGCGGTGG 386
Qy 681 hrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnH 701
Db 385 TCACCGCGTGGCCCGCAGCAGCGCGCGCGCAATACAG-----CGGTCCGCGCAGC 336
Qy 701 isGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-Val 720
Db 335 ACCAGGTGGATACGGTGGCGCGCGCCACCGCGGTGATCAATGCGCAGCAACGCGACGTG 276
Qy 721 TyrIleGlyThrAsn-----GlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738
Db 275 ATCACCAGCAGCGCGAGTGGCGGCAAGTGTCTCTCACGCGCAACGCGCAACCGCG 216
Qy 739 SerGly 740
Db 215 ATCGGC 210

```

RESULT 2
 US-09-252-991A-13656
 ; Sequence 13656, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 13656
 ; LENGTH: 8211
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-13656
 Alignment Scores:

```

Pred. No.: 2,98e-10 Length: 8211
Score: 227.50 Matches: 218
Percent Similarity: 31.30% Conservative: 90
Best Local Similarity: 22.15% Mismatches: 360
Query Match: 5.64% Indels: 319
DB: 4 Gaps: 52

US-09-917-376-3 (1-740) x US-09-252-991A-13656 (1-8211)

Qy 20 AspGlyIleValPheAsnGluGlyAla---ProGlyIleLeuTyrValArgThrAspIle 38
Db 3574 AACCGCTCGTCATCAGCGGCACACCGCGCGTGCACCGTACCTTACCGATGCC 3633
Qy 39 GlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal 58
Db 3634 GGC-----GGCAACCGATAGGCGAGGTACCGCGCGAC 3666
Qy 59 GlyTrpAsnAsnTrpGlyTyr-----AsnGly-----ValVal 69
Db 3667 GGCAGCGCAACTGGAGCTTCACGCGCGGCACGCGCGGCCCAACGCGCAGCGTATCGTC 3726
Qy 70 SerIleAlaAlaAspProIle---AsnThrAsnLysValTrpAla---Ala 84
Db 3727 GCCACGCGCCACCGACCCGACCGCGCAATACCGCGCGCAGCGCGCCACCGCGTGGACGCG 3786
Qy 85 ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu----- 100
Db 3787 GTGGCGCGCGCGCGCGTGTATCGATCGAGCAACGCGCAGCACCATCAGCGCGCACCGCG 3846
Qy 101 -----ArgSer 102
Db 3847 GAGCGCGCGGCCAAGGTGATCTCACGCGGCAACGCGCAACCGATCGCGCAACCAACC 3906
Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGly----- 119
Db 3907 GCCACGCGCAGCGCGCAACTGGACCTTCACGCGCGCACCGCGCTGGCCCAACGCGCACCGGT 3966
Qy 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArg 130
Db 3967 GTCAACGCGGTGGCCAGCACCGTCTGGGGCAATACCGGCGCGCGGCGGCGACTACCGGTG 4026
Qy 131 LeuAlaValAspProAsn-----AsnAspAsnIleLeuTyrPhe 143
Db 4027 GAGCGGTGGCGCGCGCAACACGCGCTGTGGTCAATCCGAGCAACGCGCAACCTGTCTCAACGGT 4086
Qy 144 GlyAlaProSerGlyLysGlyLeu-----Trp 152
Db 4087 ACCGCGCGAGCGCGCGCAGCACCGTGTACCGCGCAACGCGCAACCGATCGCGCCAG 4146
Qy 153 ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe 167
Db 4147 ACCACCGCGCGATGGCAGCGCGCAACTGGAGCTTCACGCGCGGTCTCGCAACTACCCCAAC--- 4203
Qy 168 ProAspValGlyThrTyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAsp 186
Db 4204 -----GGCACCGTGTCAACGTGACGCGGAGCGCGCGCGCGCGCAATACCGAC--- 4251
Qy 187 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln----- 204
Db 4252 -----GCTCCCGCTACCGACGCGGTGATTCCTCGCTCGTGTGATCCCGCAGGTGGAT 4305
Qy 205 ---AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe----- 221
Db 4306 CCGAGCACAGGTTCGGTGTATCAGCGCGCACCGCGCGCGCGCGCAACCATCATCATCATCACC 4365
Qy 222 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAla 232
Db 4366 GATGGCAACGCGCAACCCGATTTGGCCAGGTACCGCGCAGCGGCGGCGGTAACTGCTTCTTC 4425
Qy 233 ValProGlyAlaPro----- 237
Db 4425 ACTCCAGGATCCCGTCCGGATGGCACGGTGGTCAACGTGGTGGCGCGCAGCGCCCAAGC 4485

```


QY 238 -----ThrGlyPheIleProHisLysGlyVal 246
DB 4486 AATGTCGACAGTGC CGCGCGGTGATCACTGTGTGATGGCGCGCGCGCGGTG 4545
QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr--- 265
DB 4546 ATCGATCCGACCAACCGCACCGAGATAGCGGTACCGCGAGGCGCGCGCGGTGATC 4605
QY 266 -----AspGlySerSerGlyAspValTyrPheSerValThr---SerGlyThrTrp 282
DB 4606 CTCACCGATGCGCGCGCAACCCGATCGGCCCGACCGCCCGCGAGCGGCAACTGG 4665
QY 283 ThrArgIleSerProValPro-----SerThrAsp 292
DB 4666 ACGTTCAACCCCGGCACCCCGCTGGCCNACCGCACCGGTGATCAACCGCGTGGCCCGAC 4725
QY 293 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 310
DB 4726 CCGCGCGGCAATACCGAGCGGTCCGCGCAGCGTACCGTCGATGCCATCGCCCGCGCGG 4785
QY 311 -----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324
DB 4786 CCGGTGATCAATCCGACGACGAGTGTGTCATCAGCGGTACGCGGAGAGCGCGGCCACG 4845
QY 325 ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr 338
DB 4846 GTGATCTTC-----ACGACGGCNAACGGCAACCGCATCGGCCAGGTCAACCGCCAC 4896
QY 339 ArgIleTrpAspTrpThr-SerTyrProAsnArg-----SerIle 351
DB 4897 GGCAGCGCAAGTGGGCTTTCAGCGCGCCACCGCGTGGCCAAATGGCACGGTGTATCAAT 4956
QY 351 uArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPr 371
DB 4957 CGCGTGG-----CCGAGACCGCCCGCGCAACACAGCAGTCT-----CC 4995
QY 371 oProValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAs 391
DB 4996 ACCAGCGCA---CGTGCATCGGTGGCGGACGCA-GCCCGGTGATGATCGAGCA 5051
QY 391 nSerAspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAs 408
DB 5052 CGGTAGC---GTGATCGCGGTACCGCGAGCGGTGGTGCACCGGTGATCTCACC- 5103
QY 408 pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuG 428
DB 5104 -----GACGGCAACGGCAAC-----CCGATCGGCCAGGTCAACCGC 5138
QY 428 uGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu----- 444
DB 5139 CGATGGCAGCGCAACTGGAGTTCACGCC-----GGCAGCGCGGTGTCCAATGGCAC 5192
QY 445 -----IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAl 461
DB 5193 GGTGTCTCAATGGGTGGCCAGGAGCGTGGCGGCAACACCGCGCCCGCGCCAGCACCA 5252
QY 461 aValProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal-- 478
DB 5253 GGTGACTCGTGGCGCGCGCGCGTGTGATCGACCGGCAACCGCGAGCGGTGATCGC 5312
QY 479 -AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 498
DB 5313 CGGTACCGCGGAACCGGTGGCGAGCGGTGATCTCACCAGTGGCGCGCAACCCGATCGG 5372
QY 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 515
DB 5373 CCAGGCC-----ACGCGCATGGCAGCGCAACTGGAGTTCAC 5411
QY 516 -----PheGlnGlySerGluProGlyG 523
DB 5412 CCGCGGCGCGCGTGGCGCAACCGCACCGGTGATCAATCGGTGGCGCCAGGATCCGCGCG 5471
QY 523 yValThrThrGlyGlyThr----- 529

DB 5472 CAATACCGCGCGCGCGACCGACCGCGTGGAGCGCGTGGCCCGCGCCACCCCGGTGT 5531
QY 530 -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAl 542
DB 5532 CAACCGGAGCAACCGCAGCGTGTACCGGTACCGCGGAGCGCGCGCCACCGTATCTC 5591
QY 542 aProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTr 560
DB 5592 CACCGACCGCGCGCGCAACCGATCGGCCAGGTCAACCGCGCGCGCGCGCAAC---TG 5648
QY 560 pAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer----- 574
DB 5649 GAGTTCACCGCGCGCGCGCGTGGCCACCGGTGCGTGTCAATGCGTGGCGCCAGCA 5708
QY 575 -----AspArgValAsnProLysThr 581
DB 5709 CGCCCGCGCGCAACACCGCGCGCGCGCGCACCGATCGTGTGACTCGGTAGCCCGCGCC 5768
QY 581 r---PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer----- 593
DB 5769 CCGGTGTCTGATCCGAGCAACCGTGTGATCAGCGGTACCGCGCGGAGCGCGGCCAC 5828
QY 594 -----ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerS 610
DB 5829 GGTGATCTCTACCGCGCGCGC-----GGCAACCCGATACCGCGAGCGCCCGCGCATGG 5882
QY 610 erGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630
DB 5883 CAGCGCGCAACTGGAGTTCCTCCCGCGGCACA-CCGCTGACCAACCGCGCGGTGATCAATG 5941
QY 630 laAlaSerSerGlyLeuTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaI 648
DB 5942 CGGTGGCGCGAGCGCGCGCGCAACACCGCGGTCCGCTCAGCACCACAGTGGCGCGG 6001
QY 648 leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerT 668
DB 6002 TGGCGCGCGCGCACCGCGGTGATCGACCGAGCAATGGTGTCAAACCTCAGCGCGCACCGCG 6061
QY 668 yrProAlaValPheValValGlyThrIleGlyGly----- 679
DB 6062 AACCGCGGTCTCGGTGTGATCTCTACCGATGGCAATGGCAACCCGATCGGCCAGACCTCG 6121
QY 680 -----ValT 681
DB 6122 CCGAGGTAGCGGCAACTGGACCTTCACACCGCGCGCGCGTGGCCCAACCGCACCGTGG 6181
QY 681 hrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnH 701
DB 6182 TCNACCGCGTGGCGCGCGCGCGCGCGCGCAATACCG-----CGTCCGCGCGCAGC 6231
QY 701 isGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-Val 720
DB 6232 ACCAGGTGGATACCGTGGCGCGCGCGCGCGCGGTGATCAATGCCAGCAACCGCGCGGT 6291
QY 721 TyrIleGlyThrAsn-----GlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738
DB 6292 ATCACCAGCGACCGCGAGTGGCGCGCGCAAGTGTCTCACCAGCGCAACCGCAACCCG 6351
QY 739 SerGly 740
DB 6352 ATCGGC 6357

RESULT 3

US-09-252-991A-13774/c
; Sequence 13774, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 13774
 ; LENGTH: 4188
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-13774

Alignment Scores:
 Pred. No.: 2,05e-08 Length: 4188
 Score: 203.00 Matches: 186
 Percent Similarity: 32.38% Conservative: 85
 Best Local Similarity: 22.22% Mismatches: 297
 Query Match: 5.03% Indels: 274
 DB: 4 Gaps: 46

US-09-917-376-3 (1-740) x US-09-252-991A-13774 (1-4188)

```

QY 47 AlaAsnGlyArgTyr-----IleProLeuLeuAspTyrValGlyTyrAsn 61
DB 2496 GCCAACGGCAACTGCTTTCACGCGGTCCACCCCGCTGCCGACG--GTACCGTGGTCA 2439

QY 62 AsnTyr-----GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsn 79
DB 2438 ACGTGTGGCCAGGATGCGCGCGCAACAGCAGTCCGCGCCGACGTTA-----2388

QY 80 LysValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspProAsnAspGlyAlaIle 99
DB 2387 ---CGTGGAT--GCCGTGCGCGCGCCACGCCCGCTCGATCCGAGCAACCGTACGACC 2332

QY 100 LeuArg---SerSerAspGlnGlyAlaThrTyrGlnIleThrProLeuProPheLysLeu 118
DB 2331 CTCAGGCGCACCGCGCGCGCGCTACCTGACCTGACC-----2290

QY 119 GlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAsp 138
DB 2289 GACGGCAAC-----GGCAACCGGATGGCCAGGTACCGCC-----2254

QY 139 AsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGly 158
DB 2253 -----GACGGCAGCGGCAACTGG-----2236

QY 159 AlaThrTyrSerGlnMetThrAsnPheProAspValGlyThrTyrIle---AlaAsnPro 177
DB 2235 ---ACCTTACCCCGAGCAGCGCGTGTGCCCAAC---GGCAGGTGTGTCAACGCCACGGCT 2182

QY 178 ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLys 197
DB 2181 ACGACCGCTCGGCACACCGCAGTTCGCGCGCCAGCGGTACCGTGGACCCCGTGGCACCG 2122

QY 198 SerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---216
DB 2121 GCCACGCCAGTGGTCAACCCGAGCAACGCGCACCGCTCAGCGGCACCGCGGACCGCGCGGC 2062

QY 217 -----AsnAsnProValPheTyr---SerArgAspGly 226
DB 2061 GCCACCGTGACCTGGCCGATGGCAACGGCAATCCATCTCGGGCAGGTCTACCGCGGATGGC 2002

QY 227 GlyAlaThrTyrGlnAlaValProGlyAlaPro-----237
DB 2001 AGCGGCACCTGGAGCTTCACTCGACACCGCTTGTGCCCAACGGCACCGGTGTCAACGCC 1942

QY 238 -----ThrGlyPhe 240
DB 1941 ACGGCCACCGACCGCTCCGCGCAACACAGTGGCGGCGAGCGTGTACCGTGGACTCGGTA 1882

QY 241 IleProHisGlyGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
DB 1881 GCCCGCGCGCACGTCAGTGTATCAACCCGACCAACCGGACCAACCGTCTACGCGCGCAC 1822
  
```


[illegible]

RESULT 4

```

US-09-252-991A-13773/c
; Sequence 13773, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13773
; LENGTH: 4131
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13773

```

Alignment Scores:		
Pred. No.:	2.77e-08	Length:
Score:	201.50	Matches:
Percent Similarity:	31.62%	Conservative:
Best Local Similarity:	22.10%	Mismatches:
Query Match:	4.99%	Indels:
DB:	4	Gaps:
		43
		4331

US-09-917-376-3 (1-740) x US-09-252-991A-13773 (1-4131)

Qy	48	AsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr-----	65
Db	3954	AACGGCAACCGATTGGCCAGGTCACCGCCAGCGGCAAGCTGGTCTTCACTCCA	3895
Qy	66	-----AsnGlyValVal-----SerIleAlaAlaAspProIleAsnThr	78
Db	3894	GGCATCCCGTCCGGATGGACCGTGTCAACGTGTGGCGCGCACCCCAAGCAATGTC	3835
Qy	79	AsnLysValTrpAla-----AlaValGlyMetTyrThrAsnSerTrpAsp	93
Db	3834	GACAGTCGGCGCGCGTGATCACTGTGGATGGCTGGCCCCGGCGCGCGGTGATCGAT	3775
Qy	94	ProAsnAspGlyAla-----IleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr	112
Db	3774	CCGAGCAACCGCACCGAGATAAGCGGTACCGCGGAGGCGCGCGAGTCCTCACC	3715
Qy	113	ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla	132
Db	3714	GAT-----GGCGGCGGCAACCCCG-----ATCGCGCCAGGCCACCGCC	3679
Qy	133	ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp	152
Db	3678	-----GACGCGACGGCGCACTGG-----GACGCGACGGCGCACTGG	3661
Qy	153	ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr	172
Db	3660	ACGTTCAACCCCG-----GGCACC-----GGCACC	3643
Qy	173	TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp	192
Db	3642	CCGCTGCGCCAACCGCACCGTGATCAACGCC-----GACGCGACGGCGCACTGG	3613
Qy	193	ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePhe-----	210
Db	3612	GTGGCCAGGACCCGCGCGGCAATACAGCGGTCCGCGCAGCGTCACCGTCAGTGCCATC	3553
Qy	211	-----ValGlyValAlaAspProAsnAsnProValPheTrpSer-----	223
Db	3552	GCCCGCGCGCGCGTGATCAATCCGAGCAACGGAGTCGTTCATCAGCGGTACGGCGGAA	3493
Qy	224	-----Arg-----Arg	224
Db	3492	GCCGGGCGCACGGTGATCTCTCACGACGGCAACCGCAACCGGATCGCGCAGGTCAACGCC	3433
Qy	225	AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro-----	237
Db	3432	GACGCGACGGCAAGTGGCTTTTCACGCCCGCCACCGGTGGCCAAATGGCAGGTGATC	3373
Qy	238	-----Thr-----Thr	238
Db	3372	AATCGCTGCGCCAGGACCGCGCGCAACACAGCAGTCCCACACGCGCCACCGTCGAC	3313
Qy	239	GlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr	258
Db	3312	TCGCTGGCGCACGACCGCCCGGTGATCGATCCGAGCAACGGTAGCTGATCGCCGGTACC	3253
Qy	259	SerAsnThrGlyGlyProTyr-----AspGlySerSerGlyAspValTrpLysPhe	275
Db	3252	GCCGAGGCTGTGCGCACCGGTGATCCTCACCGACGGCAACCGCAACCGATCGCGCAGGTC	3193
Qy	276	SerValThr-----SerGlyThrTrpThrArgIleSerProValPro-----	289
Db	3192	ACCCCGATGGCAGCGCAACTGGAGTCTCACGCCCGCGCACCGCGCTGTCCAAATGGCAGC	3133
Qy	290	-----SerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThr	304
Db	3132	GTGTCATCGGTGGCCAGGACGCTGCCGGCAACACCGCGGCCCGCGCAGCACCCACG	3073
Qy	305	IleAspArgGlnHisPro-----AsnThrIleMetValAla	316


```
QY 94 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 113
Db 427012 AACAACTTTGGC-----
QY 114 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 133
Db 427000 -----ATGCCAACCGCGGCAACTTCAACACCGGCACTTCC 426965
QY 134 AspProAsnAsnAspAsnIleLeuTyrrPheGlyAlaProSerGlyLys--GlyLeuTrp 152
Db 426964 AACACCGGCACAAACAACATCGCTTGTTCACACCGGCACAAACAACATCGGCACTTGG 426905
QY 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPhePheAspValGlyThr 172
Db 426904 CTGACCGCGGACGGTGTCTCGGCTTCACTCCCTGAACTCCGGCGCGGCAACACCGGT 426845
QY 173 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 192
Db 426844 TTCCTCAACTCGGCACCGCCACACACCGGC----- 426815
QY 193 ValAlaPheAspLysSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal 211
Db 426814 ---TTGTTTCAACTCGGCACCGGCACACACCGGCTTGTTCACATCGGCACCGGCAACGTC 426758
QY 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231
Db 426757 GGCATCGGCACATCGGCACCGCGGCTTCGGCTCGGCCTATCCGGGACAGCAGGTG 426698
QY 232 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 251
Db 426697 GGCATCGGCACCGCACTCGGCAGTTTC---AACATCGGCTTGTTCACATCGGCACCG 426641
QY 252 HisValLeuTyrrIleAlaThrSerAsnThrGlyGlyProTyrrAspGlySerSerGly-As 271
Db 426640 GGCAATGTCGGCATCGCAACTCGGCACCGGCACCGTTCGGCATCGGCACACACCGGCACC 426581
QY 271 pValTrpLysPheSerValThrSerGlyThr-----Trp 282
Db 426580 GGCAACACCGGCATCGGAACACGCGGCAACTACACACCGGCTTCTCAACCGCGGCGCTG 426521
QY 282 p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs 296
Db 426520 GTCAACACCGGCATCGCAACCGCGGCAACACACACCGGCTTTCACACATCGGCACC 426461
QY 296 pTyrrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetVal 316
Db 426460 TTCAACACCGGCATCGCCACCGCGGCACCTACACACCGGCTCTCAACACACCGGTAGC 426401
QY 316 aThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSerThr----- 331
Db 426400 TACAACACCGGCATCGGAACACCGCGGACCTACCGCACCGGCGCTTCATCACCGGCAGC 426341
QY 332 -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer-- 345
Db 426340 ATGAACAACGGTGTCTCGGCGCGCCACCGGACGGGCGCTCTCGCGGCACACTACACC 426281
QY 346 -----TyrProAsnArgSerLeuArgTyrrValLeuAspIleSerAlaG 360
Db 426280 ATCAACATCGAGCAGCTCGCGGCTCTCAATGTCGACATCCCGGTCAACATCCCCATC 426221
QY 360 uProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTr 380
Db 426220 ACCG---GCGACATCAACATGTCTCCATCCCGGCAATTCATGTTCCCCAGA-ATC----- 426170
QY 380 pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyrrGlyTh 399
Db 426169 ---GACCGCCAGGAACGGTCGATAGGCATTCCTAGTGGCACCGGCTCTTGGCCCGCGGT 426114
QY 399 rGly---AlaThrLeuTyrrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnI 418
Db 426113 CGGTCCGATCACCTTCATCGCGGGGACGCGTCCGCGCCCGCTTGGCCCGCGGT 426054
```



```
US-09-902-540-1027/c
; Sequence 1027, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldmann, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1027
; LENGTH: 10317
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1027

Alignment Scores:
Pred. No.: 5,05e-06 Length: 10317
Score: 184.00 Matches: 155
Percent Similarity: 29.41% Conservatives: 70
Best Local Similarity: 20.28% Mismatches: 228
Query Match: 4.56% Indels: 312
DB: 4 Gaps: 40

US-09-917-376-3 (1-740) x US-09-902-540-1027 (1-10317)
QY 44 TrpAspAlaAlaAsnGlyArgTrpPro----- 53
Db 4722 TGGGATGCGGAGCGGCGGCTTCATCATGACAGCGGACGGGCTCGGGTGCACAAACCG 4663
QY 54 -----LeuLeuAspTrpValGly----- 61
Db 4662 CCCACGGGGTGTCAATGACACGGTGGGCTTCACCTCGTCGCGACCGGGACCTGGGAG 4603
QY 62 AsnTrpGlyTyraAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysVal 81
Db 4602 GAGTGGCGGTCGCGGACTCTCGGCGCGCGGCGCTCCGCGCGGTCGGTGGCGCC 4543
QY 82 TrpAlaAlaValGly----- 86
Db 4542 TTCAGATGCGGAGCGGCGCATACGGTGTCTTCGGTGGAGAGAGCTGCACGGGGTGC 4483
QY 87 MetTyraAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGly 106
Db 4482 AAGCTGTCCGACACGTGGCAGTTCAACTCAGCCACT----- 4447
QY 107 AlaThrTrpGlnIleThrProLeuProPheLeuLysLeuGlyAsnMetProGlyArgGly 126
Db 4446 CGGACTTGGCAG-----CAGGTGGGGGCGCGGGGCGCTCCGCGCT 4405
QY 127 MetGlyGluArgLeuAlaValAspProAsnAspAsnIleLeuTyraPheGlyAlaPro 146
Db 4404 TCGGGGCGGCGTGACGCGCTCCCGGTCTCGGCGCATCTTGTGTTCGAGGTACC 4345
QY 147 SerGlyLysGly-----LeuTrpArgSerThrAsp 156
Db 4344 TCGGGAGCGGCTCGCAGGAGGCGCTTTCGTCAACACGAGCTTTGGAGCCATGCCCCAG 4285
QY 157 SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyraIleAlaAsn 176
Db 4284 GGTGGGCGCAAGTTGTCTCCACCGGGAGCTTCCCGCGCGCGCGCTTCGCGCGCGCC 4225
QY 177 ProThrAspThrTrpGlyTyraGlnSerAspIleGlnGlyValValTrpValAlaPheAsp 196
Db 4224 CGCTCGACACGAGC-----TCCGGGCTCATGTGGATCTGCGCGGT 4183
QY 197 LysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro 216
Db 197 LysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro 216

4182 GCGAGCGGCGGCGCCCATCGGC-----GCCACCTG 4153
QY 217 AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAla 236
Db 4152 GCGAGTGTCTCCAGTACACCGCTCGACCATGCG-----TGGCGAGCGCGCTCGCTG 4096
QY 237 ProThr-----GlyPheIleProHisLysGlyValPheAspPro 249
Db 4095 CCGCGGCGTCCGTCGCGTGGATTGATGCTTATCGCGCGCGCGGAGTAT----- 4042
QY 250 ValAsnHisValLeuTyraIleAlaThrSerAsnThrGlyGlyProTyraAspGlySerSer 269
Db 4041 -----CTCTTCGAGGACGTCAC-----GGTGGCGCGCGC 4009
QY 270 GlyAspValTrpLysPheSerValThrSerGlyThrTrpArgIle----- 285
Db 4008 AACGATCTGCTCCGTTTCAGC-----GCGGGGCGTGGACACGCTACCGCGCAGGC 3955
QY 286 -----SerProValProSerThrAspThrAlaAsnAspTyraPheGlyTyraSerGly 302
Db 3954 GCGAGCGGCGGAGTCTCT-----CGAGCACT-----GGTGG 3925
QY 303 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp 322
Db 3924 GTGAAGCTCGAGACGGAC-----CCACGACA----- 3898
QY 323 ProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpArgIleTrpAsp 342
Db 3897 GCGGACCTCTCTCTCTCGCGGAC-----GGGCGGTGTGGAC-----TTGCAC 3850
QY 343 TrpThrSerTyraProAsnArgSerLeuArgTyraValLeuAspIleSerAlaGluProTrp 362
Db 3849 GAACCGCGTGGCAGCGGCGCACT-----TCGTCGTCTCGCGCGGAGACC 3805
QY 363 LeuThrPheGlyValGlnProAsnProValProSerProLysLeuGly----- 379
Db 3804 CTCAGCTCTCGGGGACGATTTCGCGGAGTTCAGCGCGCTTCAGTGGCGCCACCATC 3745
QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyraGlyThr 399
Db 3744 TGGGTG-----GTCGCGCCACCGCGTTCAGCGCGACCATCTTACGTC 3700
QY 400 GlyAlaThrLeuTyraAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 419
Db 3699 GGTCTGACG-----GGTGG----- 3685
QY 420 IleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProPro 439
Db 3684 -----GTCCGACGCTACAGCTCAGCGGGTGGCG 3655
QY 440 SerGlyAlaProLeu-----IleSerAlaLeuGlyAspLeuGly-----PheThrHisAla 457
Db 3654 GCGGGGTTCGCTCTCCGTCTATGCTACTACGAGTCCGAGGCGCTTATTTCTCACAAG 3595
QY 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 477
Db 3594 GAGCTGGGGAGTGGGCGCGCTGGTC-----GCGGACACGAGC 3556
QY 478 ValAspTyraAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 497
Db 3555 TTGGACATCGG----- 3544
QY 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 517
Db 3544 ----- 3544
QY 518 GlySerGluProGlyGlyVal-----ThrThrGlyGlyThrValAlaAlaSerAla 534
Db 3543 -----CTGGACACCGCGCGCGCGCGACCGACCGACCGCTCGCTCGCTCCCGCG 3487
QY 535 AspGlySerArgPheValTrpAla----- 542
Db 3486 GATTGGCGCGCTTCATCATGCGCGCTATGCGCGGAGGGTTCGCTCCAGTCAGGACTT 3427
```


Qy	543	ProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAenSerTrpAlaAla	562
Dd	3426	CCCGCAGGCCCAATGGCTCCCACAGAGCACGGCGGCACGTTCGCGC	3367
Qy	563	SerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValaenProLysThrPhe	582
Dd	3366	ACCTAC---ATCCCTGGCTCGCA--	3334
Qy	583	TyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnPro	602
Dd	3333	GCGGTCTCGCTCGAGGCACGGTC-----	3310
Qy	603	ValAlaAlaGlyLeu-----	612
Dd	3309	GTGGGGCATCGCTCTCGGAGACGTTGACAGCACTGGGTCTACAATCTCCGCGGGTGCT	3250
Qy	613	ValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer	632
Dd	3249	GTGCGCGCGTG---GCGTACCG-----	3229
Qy	633	SerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSer	652
Dd	3228	-----GATGGCATCCGGGGCGTG---	3211
Qy	653	SerAlaValaenValGlyPheGlyLysSerAlaProGlySerSerSerTyrProAlaValPhe	672
Dd	3210	-----GCGCGGGGCTGAGTGAATCGCGGCACCGCG	3181
Qy	673	ValValGlyThrIleGlyValThrGlyAlaTyr-----	684
Dd	3180	ACGGTGCGACGGTC-----GCGGAGGGGGCGTATCATCGTACC CGCGCGGGGCACG	3127
Qy	685	-----ArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAsp	699
Dd	3126	GAGTTGATCTCCGTGCGCTCGAGGCGCGGAGGATGTCGTGGGACTGGTGTCATCTCTCG	3067
Qy	700	GlnHisGlnTyrGly	704
Dd	3066	CAGGCACAGCGGCGC	3052

RESULT 7

```

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
;
GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R. A. M.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
;
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;
FILE REFERENCE: 24366-20007.00
;
CURRENT APPLICATION NUMBER: US/09/103,840A
;
CURRENT FILING DATE: 1998-06-24
;
NUMBER OF SEQ ID NOS: 2
;
SOFTWARE: PatentIn ver. 2.1
;
SEQ ID NO 2
;
LENGTH: 4403765
;
TYPE: DNA
;
ORGANISM: Mycobacterium tuberculosis
;
FEATURE:
;
OTHER INFORMATION: CDC 1551
;
OTHER INFORMATION: "n" bases at various positions throughout the sequence
;
OTHER INFORMATION: represent a, t, c or g
;
US-09-103-840A-2

```

Alignment Scores:	
Pred. No.:	0.067
Score:	184.00
Percent Similarity:	32.85%
Best Local Similarity:	20.82%
Query Match:	4.56%
Length:	4403765
Matches:	204
Conservative:	116
Mismatches:	306
Indels:	361

DB: 3 Caps: 50

US-09-917-376-3 (1-740) x US-09-103-840A-2 (1-4403765)

QY 14 GlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu--- 32
 Db 3745039 GCGGTGGCGGTCAACCGATGGC-----GATCGGGCGGTTGGCGGTGCGAATCT 3744989

QY 33 -----TyrValArgThrAspIleGly 39
 Db 3744988 GTTGGGCGAGAACGCCGACGATCGGGCGGTGAGAGTTGATGTGGGC 3744929

QY 39 ----- 39

Db 3744928 CCGCATGTGGCGCGATGCGCGCTACCATTCGGCGCGTGGCTGCTCGCGCGCGCTT 3744869

QY 40 -----GlyMetTyrArgTrpAspAlaAlaAsnGlyArg----- 50
 Db 3744868 GCGGCGTTTCAGCCACCAGCGCGAGG-CGTTGGGGGTGTGTGCGGCGCTTCTCAATG 3744810

QY 51 -----Trp-----IleProLeu 54
 Db 3744809 CTCATTTCGCGACCGCGAAGATGTTGAGGCTTAACGGCGGCTTGGCAATGTCGGTA 3744750

QY 55 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAsp 74
 Db 3744749 ATTACAACGTCGGGTTGGCAATGTCGGGATAT-----TCAACCTGGGCGCAG 3744702

QY 75 ProIleAsnThrAsnLysValTrp-----AlaIleValGlyMetTyr 88
 Db 3744701 CCAATGTCGTGGCGCAATTTGGGTGCTGCCAACGCGGTAGCGGAATTCGGTTTCG 3744642

QY 89 ThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArg----- 101
 Db 3744641 GCAATATCGGCACGCCCACTTCGGGTTCGGCAACTCGGTCTTGGGTTGCCCGCGGCA 3744582

QY 102 -----SerSerAspGlnGlyAlaThr 108
 Db 3744581 TGGGCAATATTGGGTTGGCAATCGCGGCAGCAGCAACTACGGGCTTCGAACTGGGTG 3744522

QY 109 TrpGlnIleThrProLeuProPheLysLeu-----Gly 119
 Db 3744521 TGGGCAACATCGGTTTTGCCAACACCGGTAGCAACACATCGGGATCGGTTGACCGGG 3744462

QY 120 GlyAsnMetProGly-----ArgGlyMetGlyGluArgLeuAlaValAsp 134
 Db 3744461 AC-AACCTGATGTCATTTGGGGCGCTCAATTACGAAACCGGTAACTCGGGTTGTTCAAC 3744403

QY 135 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys---GlyLeuTrpArg 153
 Db 3744402 TCGGGCACCGCAACATCGGTTCTTCAATTCCGGGACCGCAACTTCGGGGTATTCAAC 3744343

QY 154 Ser-----ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAsp 169
 Db 3744342 TCGGGCAGCTACAAACACCGGTGTCGGTAATGCGGGACGCGCCAGTACCGGGTTGTTCAAC 3744283

QY 170 Val-----GlyThrTyrIle 174
 Db 3744282 GTTGGTGGGTTCAACACGGGTGTGGCAACGTTAGCTATTAACCGGCAGCTTCAAC 3744223

QY 175 AlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 194
 Db 3744222 GCGGGCAACCAATACATCGGTGGCTTCAACCCGGGCAACGTCAACACCGGCTCGCTGAAC 3744163

QY 195 PheAspLysSerSerSerLeu-----GlyClnAlaSerLysThrIlePheVal 211
 Db 3744162 ACCGGCAACCAACACACCGCATCGGCATCGGCCAATCTCGGGCAATGTCAACACCGCGGTTTCATC 3744103

QY 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgSp----- 225
 Db 3744102 TCG---GGCAACTTCACCAACGTTGTCTGTGGGGGTACTACGAGGCGCTGTGGGGG 3744046

QY 226 -----GlyGlyAlaThrTrpGlnAlaValPro-----GlyAlaProThrGlyPhe 240

Pred. No.: 1.05e-05 Length: 11679
Score: 181.50 Matches: 204
Percent Similarity: 31.05% Conservative: 90
Best Local Similarity: 21.54% Mismatches: 344
Query Match: 4.50% Indels: 309
DB: 4 Gaps: 54

US-09-917-376-3 (1-740) x US-09-328-352-1377 (1-11679)

QY 7 ThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGly----- 21
DB 1627 ACGTGGACA--GTAAGTGTGCGGTAGTGGTTGGTGTGATGCGAGATAAACAGATT 1683

QY 22 -----IleValPheAsnGluGlyAlaProGlyIleLeuTyValArg---ThrAsp 37
DB 1684 GATGCTAAAGTAACGTTTACAGATCAGCAGGTAATACAGACACTGTTAAAGTACGCAA 1743

QY 38 IleGlyGlyMetTyArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 57
DB 1744 ATT-----TATACATTAGACACAGCT--GCTCCTGCGAGCCAGTAATCGACCCA 1791

QY 58 ValGlyTrpAsnAsnTrpGlyTyAsnGlyVal-----ValSerIleAlaAspPro 75
DB 1792 GTT-----AACGGGACAGACCCCAATTACAGGTACAGCAGAACCT 1830

QY 76 IleAsnThrAsnLysValTrpAlaAlaValGlyMetTyThrAsn----- 90
DB 1831 GGTTCAACAGTAACAGTA-----ACCTATCTTAATGGTGACACAGCAACA 1875

QY 91 -----SerTrpAsp---ProAsnAspGlyAlaIleLeuArgSer 102
DB 1876 GTTGTAGCAGACCGGACGCGAGTTGGTCAGTACCACAAACCTCGGC-----CTTAATGAT 1929

QY 103 SerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn--- 121
DB 1930 GGTGACGAAGTTGAGGCAATTGCTACAGATCCA-----GCAGGCAACCCA 1974

QY 122 --MetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsn----- 137
DB 1975 TCTTTGCCAGGTACAGCTACTGTTGAT-----GCAGTTGGTCCAAATACCGATGGTGT 2028

QY 138 -----AspAsnIleLeuTyPheGlyAlaProSerGly 148
DB 2029 AACTTTACGGTTGATTAGTAACAGCTGCAATGTGATTAAATGCATCAGAACGCTCAGCG 2088

QY 149 Lys-----GlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
DB 2089 AACGTTACTGTTACTGTTGATTGAAACCGTTCCGCGAGATGCAGCAATACAGTGGTC 2148

QY 165 ThrAsnPheProAspValGlyThrTyIleAlaAsnProThrAspThrGlyTyGln 184
DB 2149 ACTGTTGTAATCAATGGCCAGACGTATCTGCACTGATAGTAGACAGCAGCGC----- 2202

QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204
DB 2203 -----ACATGGACAGTAGCAGCTACCAGGTAGTACTTAACCTGCGGAT 2244

QY 205 AlaSerLysThrIle----- 209
DB 2245 GCAGATAGACGATTGATGCTAAAGTAACGTTTACAGATGCGGAGGTAATAGCAGCAGT 2304

QY 209 ----- 209
DB 2305 GTTAAACGATACACAAACATATACAAATCGATACCACTGCACCTGTATGCGACAGTAATTAAC 2364

QY 210 -----PheValGlyValAlaAspProAsnAsnProValPheTrp 222
DB 2365 CCGGTTAAACGGGACAGACCGGATTACAGTACGCGAGACGCTGGTTCAACAGTAACCTGTG 2424

QY 223 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro---ThrGlyPheIle 241
DB 2425 ACTTATCCAGATGGCAGTACACACACAGTGTGTCAGGACCGGATGGCACTTGGACACTA 2484

QY 242 ProHisLysGlyValPheAspProValAsnHisValLeuTyIleAlaThrSerAsnThr 261
DB 2485 CCAAAACCCAGGTTTAAATGATGGC---GATAAAGTTTACAGCAATGCTACAGATCCAGCA 2541

QY 262 GlyGlyProTyArgAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThr 281
DB 2542 GGCACACCA-----TCATTACCAGGTACA 2565

QY 282 TrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyPheGlyTySer 301
DB 2566 GCTACTGTTGATGCGAGTGGTCCAAATACCAATGCTGTTAAT-----TTCACGGTTGAT 2619

QY 302 GlyLeuThrIleAspArg----- 307
DB 2620 TCAGTAACAGCTGACAAATGTGATTAAATGCATCAGAAGCATCAGGCAACGTTACTGTTACA 2679

QY 308 -----GlnHisProAsnThrIleMetValAlaThrGlnIle 319
DB 2680 GGTGTATTGAAAAACGTTCCGCGCAGATGCAGCAATACAGTG---GTCACTGTTGTGATC 2736

QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpTrpArg 339
DB 2737 AATGCCAGACGTATAGTGAACCTGTAGTAGCAGCAGCGC-----ACATGGACA--- 2787

QY 340 IleTrpAspTrpThrSerTyProAsnArgSerLeuArgTyValLeuAspIleSerAla 359
DB 2788 -----GTAAGCGTACCAGGTAGTAGTCTGACTGCGGATGCAGATAAGACGATT 2835

QY 360 GluProTrpLeuThrPhe----- 365
DB 2836 GATGCCAAAGTAACGTTTACAGATGCAGCAGGTAAACAGCAGCTGTTAACGATACACAC 2895

QY 366 -----GlyValGlnProAsnProProValProSerProLysLeuGlyTrp 380
DB 2896 ACATATACAGTTGATACGTTGTCACCAATGCACCCGGTG----- 2934

QY 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAsp----- 393
DB 2935 CTTGATCCGATCAATGCAACAGACCCAGTGAGCGGTGAGCAGAGCCTGGTTCAACAGTG 2994

QY 394 ArgMetLeuTy-----GlyThrGlyAlaThrLeuLeuLeuLeuAsp----- 408
DB 2995 ACTGTGACTTATCTGATGGCACCACCTGCACAGTGGTAGCAGGACCGATGTTAGTGG 3054

QY 409 -----LeuThrLysTrpAsp 413
DB 3055 TCAGTACCAAAACCCAGGTAACCTGGTGGATGGGATACAGTACAGTCAACAGCAACTGAC 3114

QY 414 SerGlyGlyGlnIleHisIle-----AlaPro 422
DB 3115 CCTGCAGGCAACACTTTCATGTCAGGTACAGGCACAGCTTTCAGCAGACATCAGACACCT 3174

QY 423 MetValLysGlyLeuGluGluThrAlaValAsnAsp-----Leu 435
DB 3175 GTGGTT---GCGCTGGATGACGTTGTGACGAATGACAGCAGCAGCACTTACGGGTACA 3231

QY 436 IleSerProProSerGlyAlaProLeuIleSerAla----- 447
DB 3232 GTGAACGATCCGACACCCACTGTAGTTGTCAATGTGGATGGCGTTGACTATCCGCGAGTG 3291

QY 448 ---LeuGlyAspLeuGlyPheThrHisAlaAsp-----ValThrAla 461
DB 3292 AACAAATGGTGAC---GGCACCTGGACGCTTGACAGACAAATACACTCTCTCGGTTAACTGAT 3348

QY 462 ValProSerThrIlePheThrSerProValPheThrThrGly----- 475
DB 3349 GGTCCACACACCACTTACCCTGACTGCAACAGATGCAGCAGGCAATCAGGTACAGATACA 3408

QY 476 -----ThrSerValAspTyAlaGluLeuAsnProSerIleIleValArgAlaGlySer 493
DB 3409 GCAGTGGTGCAGGATTGATACACAGCACCAAATGCACCGGTACTTGTATCCGATCAATGCG 3468

QY 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyLys 513

OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17474

Alignment Scores:
Pred. No.: 0.000121 Length: 29927
Score: 177.00 Matches: 189
Percent Similarity: 28.82% Conservative: 56
Best Local Similarity: 22.24% Mismatches: 292
Query Match: 4.39% Indels: 314
DB: 4 Gaps: 48

US-09-917-376-3 (1-740) x US-09-949-016-17474 (1-29927)

QY 14 GlyGlyGlyPheValaspGly-----IleValPheAsnGluGly 27
DB 19367 GGGGGCTCTGGGAGCGGAGCGCTGCCCTCCTTCCACCGTTTTATTCAAGGGGAC 19426
QY 28 AlaProGlyIleLeuTyrValArgThr-----AspIleGlyGlyMet 41
DB 19427 AGGCTGGGGATTTGTTATTGGGCGCTGTTGGCTGAGGGTGACGGACTT-GGGGGG--- 19482
QY 42 TyrArgTyrAspAlaAlaAsn----- 48
DB 19483 TGGCGGTGGGAGCGGAGGTATAAAGTATATATCATATAGTAAACAACTCAGAAATG 19542
QY 49 -----GlyArgTyrIleProLeu-LeuAspTyrValGlyTyrAsnAsnTyrGlyTyrAs 66
DB 19543 GACCCCGAGCGGTGGTGGCGGTAGCTCTCCAGCTCTCCCTGGCCCGAGGAGGAGA 19602
QY 66 nglyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTyrAlaAlaValGI 86
DB 19603 GGGGTCGGCATCCCTCCGGGTTCTCTCTC----- 19633
QY 86 yMetTyrThrAsnSerTyrAspProAsnAspGlyAla-----IleLeuArgSerSe 103
DB 19634 -----CTGGGTACCTGGCTTGGGTGGGAGACGACCTACTCTTGTACCGTCTTT 19686
QY 103 rAspGlnGlyAlaThr-----TyrGlnI 111
DB 19687 TGCCGAGCGGGGACCCAGTAAATTAGCCGTTGGAGCCCGCAGCGCTCGCTGGCTTGT 19746
QY 111 eThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGly----- 128
DB 19747 CGCACCGAGTCTTGG-----GGACCTGGTGTCTCCCGGAGAACTTGGGACCTGGTA 19800
QY 129 -----GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr----- 142
DB 19801 TCCCCGGGAGAGGCTTGGGACCTGGTGTCTCCCGGAGAGGCTTGGTACTGTTTCTCT 19860
QY 143 -----PheGlyAlaProSerGlyLysGlyLeu----- 151
DB 19861 GGAAGAGGCTTGACACCTGGTCTCTGGGAGGCGCTTGGGACCTGGTGTCTCTGGGAGA 19920
QY 152 -----TyrArgSerThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProaspva 170
DB 19921 GGCTTGAGATCTGTGTCTCTGGGAGAGGCTTGGGGA----- 19957
QY 170 lGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVa 190
DB 19958 -----CTGGGTCTCCCTGGAGA 19974
QY 190 lValTyrValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePh 210
DB 19975 GGCTTGG-----GGACCTGGTACCTTGGAGAGGCTTGGAGACCTGGTGT 20019
QY 210 eValGlyValAlaAspProAsnAsnPro-ValPheTyrSerArg----- 224
DB 20020 TCTGGGAGAGGCT-----TGGGAGCCTGGTGTCTGGGAGAGGCTTGGGAGCCTGGTGT 20073
QY 225 -----AspGlyGlyAlaThrTyr-----GlnAlaValProG 235
DB 20074 CTCTGGAAGAGGCTTGGACACCTGGTGACCCGGAGGGGCTTGGGGATCTGGGTCTCCCG 20133

QY 235 lyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuT 255
DB 20134 GAGAGCCTTGGGGA-----CTGGTGTCTCTG----- 20159
QY 255 yrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysP 275
DB 20160 -----GGAGAGGCTTGGGACCT-----GGTGAACCTTGGAGAGGCTTGGGAC 20202
QY 275 heSerValThrSerGlyThrThrArgIleSerProValProSerThrAspThrAlaA 295
DB 20203 CTGGGTCTCTGAGAGAGCCTTGGGA-----TCTGGTGTCTCCAGGAGAGGCTTGGGAC 20256
QY 295 snAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetV 315
DB 20257 CT-----GGTGTCTCTGGAAGAGGCTTGGAC----- 20282
QY 315 alaAlaThrGlnIleSerTyrTrpPro-----AspThrIleIlePheArgS 330
DB 20283 -----ACCTGGTGTCTCTGGGAGAGGCTTGGGACCTGGTG----- 20318
QY 330 erThrAspGlyGlyAlaThrThrArgIleTyrAsp-TyrThrSerTyrProAsnArg 349
DB 20319 -----TCTGGGAGAGGCTTGGGACCTGGTGT-----CTTGGGAGA 20355
QY 350 SerLeu-ArgTyrValLeuAspIleSerAlaGluProTyrLeuThrPheGlyValGlnPr 369
DB 20356 GCGTTGGAGA-----TCTGGTGGAGCGGAGAGGCTTGGGGA----- 20392
QY 369 oAsnProProValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspPr 389
DB 20393 -----CTGGTGTCTCCCGGAGAGGCTTGGGACCTGGTGTCTCCCGGAGAGGCTTGGACACC 20448
QY 389 o-PheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspL 409
DB 20449 TGGTGTCTCCAGGAGAGGCTT-----GGGACCTGGTGTCTGGAGAGGCTGGGAC 20502
QY 409 euThrLysTyrAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluG 429
DB 20503 TGTGACCCGGGAGAGGCTTGGG----- 20525
QY 429 luThrAlaValAsnAspLeuIleSer-----ProProSerGlyAlaProLeuIleSerA 447
DB 20526 -----GACCTGTGTCTCTGGGAGAGGCTTGGGACCT----- 20558
QY 447 laLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIleP 467
DB 20559 -----GGTGAACCTTGGAGAGGCTTGGGACCTTGTGTCTCGGGAGTGCCTTGGGACCT- 20612
QY 467 heThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerI 487
DB 20613 --AGTGACCCGGGAGAGGCTTGGGACCTGGTGTCTCCGGGAGAGGCTTGGGACCTGGTG 20670
QY 487 leIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaP 507
DB 20671 TCCTG-----GGAGAGCCTTGGGATCTGTGTCTCTG----- 20702
QY 507 heSerThrAspGlyGlyLysAsnTyrPheGlnGlySerGluProGlyGlyValThrThrG 527
DB 20703 -----GGGAGAGGCTTGGGACCTTGTGTCTCTCGGAGAGAGGCTTGGGACCTGGTGACC 20757
QY 527 lyGlyThrVal---AlaAlaSerAlaAspGlySerArgPheValTyrAlaProGlyAspP 546
DB 20758 CGGAGAGGCTTGGACACCTGGTGTCTCCGGGAGAG-----GCTTGGGACCTGGTGACC 20811
QY 546 roGlyGln-----ProValValTyrAlaValGlyPheGlyAsnSerTyrAlaAlaAs 563
DB 20812 CGGAGAGCCTTGGGAGCCTTGTGTCTCTGGGAGAGGCTGGGGG---ACCTGTGTCTCG 20868
QY 563 erGln-----GlyValProAlaAsnAlaGlnIleArgSerAspArgValA 578
DB 20869 GGAGAGAGCCTTGGGACCTTGTGTACCCGGGAGAGGCTTGG----- 20909
QY 578 snProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly---- 596

Db	20910	-----ACACCTGGTGTC 20922	Query Match: 4.39%	Indels: 314
Qy	597	--GlyValThrPheGlnProValAlaAlaGlyLeuPro---SerSerGlyAlaValGlyV 615	Gaps: 48	
Db	20923	CGGGAGAGGCTTGGGAGGCTGTGTCTCCGGGAGAGCCTTGGGGACACGCTGACCTTGGAG 20982		
Qy	615	alMetPheHisAlaValProGlyLysGlyValAspLeuTriPheAlaAlaSerSerGlyL 635		
Db	20983	AGGCTTGGGA-----CCTGGTGATCTGGAGGCTTGGGGACCTGGTGCTCGGA- 21035		
Qy	635	eUyThrAsnGlyGlySerSerThrAlaIleThrGlyValSerSerAlav 655		
Db	21036	-----GAGGTTACGGGGCTGGTTGGGGGAGAGAACGTTGTGTGAGCCAAAGTCCC 21084		
Qy	655	alaSn----- 656		
Db	21085	TGAATCCCTCGAAGAGAGCGCATCGGGAGTCCCCCTGAGGGCGTTCATTTGTGGACC 21144		
Qy	657	-----ValGlyPheGlyLysSerAlaProGlySerSerTyProAlav 671		
Db	21145	CCCTCCCATGGCTTTCAGGAGCTGTTCGATTCCTGGCCGGCT---CCGGCGG 21201		
Qy	671	alPheValGlyThrIleGlyValThrGlyAlaTyArgSerAspAspCysGlyT 691		
Db	21202	ATGCATCCAGTGGCAGCGCAATTCTGGGCCAGGGGAGGAGAAAGCGGGTGTGGGG 21261		
Qy	691	hrThrTrpValLeuIleAsnAspAspGlnHisGlnTyGlyAsnTrpGlyGlnAlaIle 711		
Db	21262	TGGTC-----TCCACGG 21273		
Qy	711	hrGlyAspHisAlaAsnLeuArgArgValTyIleGlyThrAsnGlyArgGlyIleValT 731		
Db	21274	CTGGAGAGGGCGAGCTCCCTAGG-----GGAGAGAGGCACGCTTGGGGTTTC 21324		
Qy	731	yrGlyAspIleGlyAla 737		
Db	21325	CGGGGGCGGGCGGAGCA 21344		
RESULT 11				
US-09-949-016-17475				
; Sequence 17475, Application US/09949016				
; Patent No. 6812339				
; GENERAL INFORMATION:				
; APPLICANT: VENTER, J. Craig et al.				
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED				
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF				
; FILE REFERENCE: CL001307				
; CURRENT APPLICATION NUMBER: US/09/949,016				
; CURRENT FILING DATE: 2000-04-14				
; PRIOR APPLICATION NUMBER: 60/241,755				
; PRIOR FILING DATE: 2000-10-20				
; PRIOR APPLICATION NUMBER: 60/237,768				
; PRIOR FILING DATE: 2000-10-03				
; PRIOR APPLICATION NUMBER: 60/231,498				
; PRIOR FILING DATE: 2000-09-08				
; NUMBER OF SEQ ID NOS: 207012				
; SOFTWARE: FastSeq for Windows Version 4.0				
; SEQ ID NO 17475				
; LENGTH: 29927				
; TYPE: DNA				
; ORGANISM: Human				
; FEATURE:				
; NAME/KEY: misc.feature				
; LOCATION: (1)...(29927)				
; OTHER INFORMATION: n = A,T,C or G				
US-09-949-016-17475				
Alignment Scores:				
Pred. No.:	0.000121	Length:	29927	
Score:	177.00	Matches:	189	
Percent Similarity:	28.82%	Conservative:	56	
Best Local Similarity:	22.24%	Mismatches:	292	

QY	275	heserValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaA	295	Db	20983	AGGCTTGGGGA-----CCTGTCATCTTGGAGAGGCTTGGGACCTGCTCTCGGA-	21035
Db	20203	CTGTGTCTCAGAGAGCCTTGGGA-----TCTGGTGTCCAGGAGAGGCTTGGGGAC	20256	QY	635	euTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlav	655
QY	295	snAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetV	315	Db	21036	-----GAGTTACGGGGCTGCTTGGGGGAGAGACGTTGTGAGCCAAAGTCCC	21084
Db	20257	CT-----GGTGTCTCTGGAAGAGGCTTGGAC-----	20282	QY	655	alaSn-----	656
QY	315	alAlaThrGlnIleSerTrpTrpPro-----AspThrIleIlePheArgS	330	Db	21085	TGAATCCCTCGGAAGAGAGCGCATCGGGAGCTCCCTCGAGGGCGTTCCATTGTGGACC	21144
Db	20283	-----ACCTGGTGTCTCGGAGAGGCTTGGGACCTGGTG-----	20318	QY	657	-----ValGlyPheGlyLysSerAlaProGlySerSerTrpProAlav	671
QY	330	erThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp-TipThrSerTyrProAsnArg	349	Db	21145	CCCTCCCATGCGCTTTGCAGGGAGCTTCCGATTCCCTTGGCCCGGCT---CCCGCGG	21201
Db	20319	-----TCTGGGAGAGGCTTTGGGACCTGGTGT-----CCTGGGAGA	20355	QY	671	alPheValValGlyThrIleGlyValThrGlyValaTyrArgSerAspCysGlyT	691
QY	350	SerLeu-ArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPr	369	Db	21202	ATGCATCCAGTGGCAGCCCAATCTTGGGCCAGGGGAGGAGAAAGCGGCTGTGGGG	21261
Db	20356	GGCTTGGAGA-----TCTGGTGAGCCGGGAGAGGCTTTGGGGA	20392	QY	691	hrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleT	711
QY	369	oAsnProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPr	389	Db	21262	TGGTC-----	21273
Db	20393	-----CCTGGTGTCCCGGAGAGGCTTGGGACTTGGTGCCCGGAGAGGCTTGGACACC	20448	QY	711	hrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValT	731
QY	389	o-PheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspL	409	Db	21274	CTGGAGAGAGGGCGGACGCTCCCTAGG-----GGAGAAGAGGCACGTTGGGGGTTTC	21324
Db	20449	TGGTGTCCAGGAGAGGCTT-----GGGACCTGGTGACCTTGGAGAGGCTTGGGACC	20502	QY	731	yrGlyAspIleGlyAla 737	
QY	409	euThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluG	429	Db	21325	CGGGGGCGGGCGGAGCA 21344	
Db	20503	TGGTGACCCGGGAGAGCCTTGGG-----	20525	RESULT 12			
QY	429	luThrAlaValAsnAspLeuIleSer-----ProProSerGlyAlaProLeuIleSerA	447	US-09-949-016-12639			
Db	20526	-----GACCTGTGTCTCGGAGAGGCTTTGGGACCT-----	20558	; Sequence 12639, Application US/09949016			
QY	447	laLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSerThrIleP	467	; Patent No. 6812339			
Db	20559	-----GGTGTCTCGAAGAGGCTTTGGGACCTGGTGCTCCCGGAGAGGCTTTGGGACCT-	20612	; GENERAL INFORMATION:			
QY	467	heThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerI	487	; APPLICANT: VENTER, J. Craig et al.			
Db	20613	-----AGTGACCCGGGAGAGGCTTTGGGACCTGGTGCTCCCGGAGAGGCTTTGGGACCTGGTG	20670	; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED			
QY	487	lelleValargAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaP	507	; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF			
Db	20671	TCTCG-----GGAGAGCCTTTGGGATCTGGTGCTCTG-----	20702	; FILE REFERENCE: CLO01307			
QY	507	heThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrG	527	; CURRENT APPLICATION NUMBER: US/09/949,016			
Db	20703	-----GGGAGAGGCTTTGGGACCTGGTGCTCTCGGAGAGAGGCTTTGGGACCTGGTGACC	20757	; CURRENT FILING DATE: 2000-04-14			
QY	527	lyGlyThrVal-----AlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspP	546	; PRIOR APPLICATION NUMBER: 60/241,755			
Db	20758	CGGAGAGGCTTTGGACACCTGTGTCTCCCGGAGAG-----GCTTGGGGACCTGGTGACC	20811	; PRIOR FILING DATE: 2000-10-20			
QY	546	roGlyGln-----ProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlas	563	; PRIOR APPLICATION NUMBER: 60/237,768			
Db	20812	CGGAGAGGCTTTGGGACCTGGTGCTCTCGGAGAGGCTTTGGGG-----ACCTGTGTCTCG	20868	; PRIOR FILING DATE: 2000-10-03			
QY	563	erGln-----GlyValProAlaAsnAlaGlnIleArgSerAspArgValA	578	; PRIOR APPLICATION NUMBER: 60/231,498			
Db	20869	GGAGAGAGCCTTTGGGACCTGGTGACCTCGGAGAGGCTTTGG-----	20909	; PRIOR FILING DATE: 2000-09-08			
QY	578	snProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly----	596	; NUMBER OF SEQ ID NOS: 207012			
Db	20910	-----ACACCTGGTGCTC-----	20922	; SOFTWARE: FastSeq for Windows Version 4.0			
QY	597	--GlyValThrPheGlnProValAlaAlaGlyLeuPro---SerSerGlyAlaValGlyV	615	; SEQ ID NO 12639			
Db	20923	CGGAGAGGCTTTGGGAGCTTTGGTGCTCCCGGAGAGGCTTTGGGACCTTTGGAG	20982	; LENGTH: 37802			
QY	615	alMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaIleSerSerGlyL	635	; TYPE: DNA			
				; ORGANISM: Human			
				; FEATURE:			
				; NAME/KEY: misc.feature			
				; LOCATION: (1)...(37802)			
				; OTHER INFORMATION: n = A,T,C or G			
				US-09-949-016-12639			
				Alignment Scores:			
				Pred. No.: 0.000175 Length: 37802			
				Score: 177.00 Matches: 189			
				Percent Similarity: 28.82% Conservative: 56			
				Best Local Similarity: 22.24% Mismatches: 292			
				Query Match: 4.39% Indels: 314			
				DB: 4 Gaps: 48			
				US-09-917-376-3 (1-740) x US-09-949-016-12639 (1-37802)			
QY	14	GlyGlyGlyGlyPheValAspGly-----IleValPheAsnGlyGly 27					
Db	19367	GGGGGCTCTGGCAGGGCGGACGCGTGCCCTCCCTCTTCCACGTTTATTCCAGGGGAC 19426					

QY 28 AlaProGlyLeuTyrValArgThr-----AspIleGlyMet 41
Db 19427 AGGCTGGGATGTAATTGGCGGCTGTGGCTGAGGGTCCAGGACTT-GGGGG- 19482
QY 42 TyrArgTrpAspAlaAlaAsn----- 48
Db 19483 TGGCGGTGGGAGCGCGAAGGTATAACGTATAATCATAGTAAACAACTCAGAAATG 19542
QY 49 -----GlyArgTrpIleProLeu-LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAs 66
Db 19543 GACCCCGAGCGCTGTCGCCCTAGCTCTCCAGCTCTCCCTGGCCCGAGGAGGAGA 19602
QY 66 nGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValG1 86
Db 19603 GGGTCCGCATCCCTCCCGGTTCTCTCTC----- 19633
QY 86 yMetTyrThrAsnSerTrpAspProAsnAspGlyAla-----IleLeuArgSerSe 103
Db 19634 -----CTGGGTACCTGGCCTTGAGGTGGGGAGAGGACCTACTTCTGTACCGTCTTT 19686
QY 103 rAspGlnGlyAlaThr-----TrpGlnI1 111
Db 19687 TGCCGACGCGCGGACCCAGTGAATTAGCCGCTTGGAGCCGCGAGGCTGCTGCTGTG 19746
QY 111 eThrProLeuProPheLysLeuGlyGlyAsnMetProGlyVArgGlyMetGly----- 128
Db 19747 CGCACCGAGGCTTGG-----GGACCTGGGTCTCCCGGAAACAACTTGGGAGCTGTGA 19800
QY 129 -----GluArgLeuAlaValAspProAsnAspAsnIleLeuTyr----- 142
Db 19801 TCCCGGGAGAGGCTTGGGACCTGGTCTCCCGGAGAGGCTTGGGTACCTGGTTCTCT 19860
QY 143 -----PheGlyAlaProSerGlyLysGlyLeu----- 151
Db 19861 GGAAGAGGCTTGGACACCTGGTGTCTCGGAGGCGCTTGGGACCTGGTCTCTGGGAGA 19920
QY 152 ---TrpArgSerThrAspSerGlyAlaThrTrpSerClnMetThrAsnPheProAspVa 170
Db 19921 GGCTTGGAGATCTGTGTCTCGGGAGAGGCTTGGGGA----- 19957
QY 170 lGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVa 190
Db 19958 -----GGACCTGGTGTGACCTTGGAGAGGCTTGGAGACCTGGTGT 20019
QY 190 lValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePh 210
Db 19975 GGCTTGG-----GGACCTGGTGTGACCTTGGAGAGGCTTGGAGACCTGGTGT 20019
QY 210 eValGlyValAlaAspProAsnAsnPro-ValPheTrpSerArg----- 224
Db 20020 TCTGGGAGAGGCT-----TGGGACCTGGGTGTCTTGGGAGAGGCTTGGGACCTGGTGT 20073
QY 225 -----AspGlyGlyAlaThrTrp---GlnAlaValProG 235
Db 20074 CTCTGGAAGAGGCTTGGACACCTGTGTACCCGGAGGCGCTTGGGATCTGGTGTCCCGG 20133
QY 235 lyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuT 255
Db 20134 GAGAGCCTTGGGA-----CCTGGTGTCTCTG----- 20159
QY 255 yrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysP 275
Db 20160 -----GGAGAGGCTTGGGACCT-----GGTGACCTTGGAGAGGCTTGGGAC 20202
QY 275 heSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaA 295
Db 20203 CTGGTGTCTGAGAGAGCCTTGGGA-----TCTGGTGTCCAGAGAGGCTTGGGAC 20256
QY 295 snAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetV 315
Db 20257 CT-----GGTCTCTCTGGAGAGGCTTGGAC----- 20282

QY 315 alAlaThrGlnIleSerTrpTrpPro-----AspThrIleIlePheArgS 330
Db 20283 -----ACCTGGTGTCTCTGGGAGAGGCTTGGGACCTGGTG----- 20318
QY 330 erThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp-TrpThrSerTyrProAsnArg 349
Db 20319 -----TCCTGGGAGAGGCTTGGGACCTGGTGT-----CCTGGGAGA 20355
QY 350 SerLeu-ArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPr 369
Db 20356 GGCTTGGAGA-----TCTGGTGAGCGGGAGAGGCTTGGGGA----- 20392
QY 369 oAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPr 389
Db 20393 ---CCTGGTGTCCCGGAGAGGCTTGGGACTTGGTGTCCCGGAGAGGCTTGGACAC 20448
QY 399 o-PheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspL 409
Db 20449 TGGTGTCCAGAGAGGCTT-----GGGACCTGGTGTACCTTGGAGAGGCTTGGGAC 20502
QY 409 euThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluG 429
Db 20503 TGGTACCCCGGAGAGCCTTGGG----- 20525
QY 429 luThrAlaValAsnAspLeuIleSer-----ProProSerGlyAlaProLeuIleSerA 447
Db 20526 -----GACCTGGTGTCTCTGGGAGAGCCTTGGGACCT----- 20558
QY 447 laLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIleP 467
Db 20559 ---GGTGACCTTGGAGAGGCTTGGGACCTGGTGTCTCTCGGAGAGTGCCTTGGGAC 20612
QY 467 heThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerI 487
Db 20613 ---AGTGACCCGGAGAGGCTTGGGACCTGGTGTCTCCCGGAGAGGCTTGGGACCTGGTG 20670
QY 487 leIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaP 507
Db 20671 TCCTG-----GGAGAGCCTTGGGATCTGGTGTCTCTG----- 20702
QY 507 heSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrG 527
Db 20703 ---GGGAGAGGCTTGGGACCTGGTGTCTCGGAGAGAGCCTTGGGACCTTGGTGAC 20757
QY 527 lyGlyThrVal---AlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspP 546
Db 20758 CGGAGAGGCTTGGACACCTGGTGTCTCCCGGAGAG-----GCTTGGGACCTTGGTGAC 20811
QY 546 roGlyGln-----ProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaS 563
Db 20812 CGGAGAGCCTTGGGACCTGGTGTCTCTGGGAGAGGCTTGGGG-----ACCTGGTGTCTCG 20868
QY 563 erGln-----GlyValProAlaAsnAlaGlnIleArgSerAspArgVala 578
Db 20869 GGAGAGAGCCTTGGGACCTGGTGTACCCCGGAGAGGCTTGG----- 20909
QY 578 snProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly----- 596
Db 20910 -----ACACCTGGGTGTC 20922
QY 597 --GlyValThrPheGlnProValAlaAlaGlyLeuPro---SerSerGlyAlaValGlyV 615
Db 20923 CGGAGAGGCTTGGGAGCCTGGTGTCTCCCGGAGAGCCTTGGGAGACAGGTGACCTTGGAG 20982
QY 615 alMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaIaSerSerGlyL 635
Db 20983 AGGCTTGGGA-----CCTGGTGTCTTGGAGAGGCTTGGGACCTTGGTGTCTCGGA- 21035
QY 635 euTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaV 655
Db 21036 -----GAGTTACGGGGCTGGTGGGGAGAGACGTTGTGTGAGCCAAAGTCCC 21084
QY 655 alAsn----- 656


```
Db 21085 TGAATCCCTGCGAAGAGCGCATCGGAGCTCCCTCGAGGGGTTTCCATTGTGGACC 21144
Qy 657 -----ValGlyPheGlyLysSerAlaProGlySerSerTyrProAlav 671
Db 21145 CCCCTCCCATCGCTTTCGAGGAGCTGTTCGATTCCTCCCTGGCCGGCT---CCCGCG 21201
Qy 671 alPheValValGlyThrileGlyValThrGlyAlaTyrArgSerAspAspCysGlyT 691
Db 21202 ATGCATCCAGTGGCAGCGCCCAATTCCTGGCCAGGGGAGGAGGAGCGCGGTGTGGG 21261
Qy 691 hrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleT 711
Db 21262 TGGTC-----TCCACGG 21273
Qy 711 hrGlyAspHisAlaAsnLeuArgValTyrileGlyThrAsnGlyArgGlyIleValT 731
Db 21274 CTGGAGAGGGGCGACGCTCCTAGG-----GGAGAAGAGCAGCTTGGGGGTTTC 21324
Qy 731 yrGlyAspIleGlyAla 737
Db 21325 CGGGGGCGGGCGGAGCA 21344
```

RESULT 13

```
US-09-902-540-2590
; Sequence 2590, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 2590
; LENGTH: 4287
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-2590
```

```
Alignment Scores:
Pred. No.: 7,98e-06 Length: 4287
Score: 175.50 Matches: 165
Percent Similarity: 35.27% Conservative: 90
Best Local Similarity: 22.82% Mismatches: 240
Query Match: 4.35% Indels: 229
DB: 4 Gaps: 47
```

US-09-917-376-3 (1-740) x US-09-902-540-2590 (1-4287)

```
Qy 24 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg 43
Db 2299 TACAACCTTCAGTTCCCGGGCATGCTCCTCGGACCATCGACTGGGACCGGCTACGCC 2358
Qy 44 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAsp-----TrpValGly 59
Db 2359 GGCAAGACGGTGCAGTCCGCTTCGGGTCGGTACGACGAGCGGGCGGCTACACGGGC 2418
Qy 60 Trp-----AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 77
Db 2419 TGCTGCTGGACGACCTCGAGTTCAACGGCATC----- 2451
Qy 78 ThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGly 97
Db 2452 ACGACACTCCGTTCCGCCACCATCGCG-----CCCGAGGATGCG 2490
Qy 98 AlaileLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLys 117
```

```
Db 2491 GTTTGT-----GTGAACCCCTGGGCCCATCGCC 2517
Qy 118 LeuGlyGlyAsnMetProGlyArgGlyMet-----GlyGluArgLeu----- 131
Db 2518 AACGCGGT-----CCGACCGGTCCATCGCGCGGTGACTGTGTAGACCTCTACGGT 2571
Qy 132 ---AlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 150
Db 2572 TCGGCTCGCATCCGAAAGCCAGCGGTGACCTTC----- 2607
Qy 151 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 170
Db 2608 -----ACCTGGGCACAG---ACGTCTGCGCTCGCGTC 2637
Qy 171 -----GlyThrTyrIleAlaAsnProThrAspThrGlyTyrGlnSerAspIle 187
Db 2638 ACCCTGGCGGGTGCACACCTGTAACCGCTGTTACGCGCGCGGAGGTATCCGAGTCC 2697
Qy 188 GlnGlyValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLys 207
Db 2698 ACCGCACTGGTGTTCACGGTG-----ACCGTCTCGACGGGTCAAGACGTCCACGGAC 2751
Qy 208 ThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGly 227
Db 2752 TCGGTGACCGTACGGTGGCGTCCCAACAATCCG----- 2787
Qy 228 AlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisGlyValPhe 247
Db 2788 -----CCCACCGTGAACCGGCGCTCGACGGCATCTCGAGAGCGCGGTGAGTAC 2838
Qy 248 AspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGly 267
Db 2839 -----ACGCTCAGCGGTCCGCCACGAGTCCGATGGC 2871
Qy 268 SerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIleSerPro 287
Db 2872 AATGCC-----CTCACGTACTGTGGACCCAGGTCTCCGGT 2907
Qy 288 ValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArg 307
Db 2908 ACGCGCGTG-----GCGGTGAAGGACTAC----- 2931
Qy 308 GlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIle 327
Db 2932 ACCACGCGACGCGGACGTTTCATCGCGCGGAGGTCAAG---CTGGATGAGTCCGCTGTC 2988
Qy 328 PheArgSerThr-----AspGlyGlyAlaThrTrp---ThrArgIleTrpAspTrpThrSe 345
Db 2989 TTTCCGTCTCAGCGTCAGCGACGCGCATCGCCACCGGTGAACG-----ACACG 3033
Qy 345 rTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPh 365
Db 3034 GTCACCGTGCAGCGTCCACCAACGCGA-----ACCGCGCGCCCA----- 3070
Qy 365 eGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAlaMe 385
Db 3071 -----TCGTCTCCGACACGTCGCGGCGCTTCGCGCGGCGACT-GTCAACCGTAGCGGCTC 3125
Qy 385 tAlaileAspProPheAsnSerAspArgMetLeuTyrGly-----ThrGlyAlaTh 402
Db 3126 CGGGTTCGACCGG---GATGGTACCGCGTCAAGTACAGTGGGAGCAGACGGGTGTC 3182
Qy 402 rLeuTyrAlaThrAsn----- 407
Db 3183 GACGGTCCGCATCAATGTGGCGCGACACGTCCGCCATCTCTTCGCCACCCCGGTCGCCGG 3242
Qy 408 -----As 408
Db 3243 TTCTGACGAGTTCCCGTGAACGCGACGAGCGGTGGCGCTCTGCCCTCCAGGGCGGTGCC 3302
Qy 408 PLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys---GlyLe 427
Db 3303 GGTGACATCATCGATGTTCTCTCGCGGCGAAGTCTCTGCGCGGCACTCTGCGCGGCGCAT 3362
```



```
Qy 427 uGluGluThrAla---ValaAspLeuIleSerProProSerGlyAlaProLeuIleSe 446
Db 3363 CGATGCCAGCGCAACGCTGGTGGACGTCGTGACGCTACGCGCTCCGCGAGC-----GA 3416
Qy 446 rAlaLeuGlyAsp-----LeuGlyGly-----Ph 454
Db 3417 CGCCGAGGGTGACACGCTGACGTACCACTGGGAGCAGATCGTGGCAGGATGTGACCT 3476
Qy 454 eThrHiAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrTh 474
Db 3477 GACGGGCGCCGAC---ACCCTGACGCGCTCG-----TTTACCGCGCGGACGACGCGCAG 3527
Qy 474 xGlyThrSerValAspTyrAlaGluLeuAsnProSer-----486
Db 3528 CGGTACACGCTCGGCTCATC---CTGACGGTCAGCGATGGCACCTCGACACACGAGCA 3584
Qy 487 -----IleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspAr 503
Db 3585 TGCCGTGCGCTCATCGTGGCGG-----GACCGGGTGCCACCCGAGC-----3630
Qy 503 gHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGl 523
Db 3631 -AACACTGCTCCGGAAGCGACGCGCGCGAGTCGCGCATCGTG---GCGAGGGCGCCAC 3686
Qy 523 yValThrThrGlyThrValAlaAlaSerAlaAspGlySerArg-----PheValTr 541
Db 3687 TGTGACGCTCAACGGCAG---GCGACGACGCTGACCGGTGACGCTGCTCATCGTCTG 3743
Qy 541 pAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAl 561
Db 3744 GACGAG---ATCGCGGCGACGCGCTG-----ACGTGTCTGACGGCTCTCGCTGAC 3794
Qy 561 aAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAlaProLysTh 581
Db 3795 GCGGAGCTTACCGCGCGCGCTCTCTCC-----GATCGGCTGAC 3833
Qy 581 rPheTyrAlaLeu---SerAsnGly-----ThrPheTyrArgSerThrAspGlyGlyVa 598
Db 3834 CTTCCTCTCGATGTGTACGACGAGTACGCGGACACTGTCTGTCGACGTCCATCAGCGT 3893
Qy 598 lThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHi 618
Db 3894 GACGAGAGAACGTGCGG-----CCGGTGGCGTCCGCTCGCGCGGTGCTC-----3939
Qy 618 sAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisE 638
Db 3940 -----TCTGGCAACGACGCTCGGCCACGCTCGACGCTCGGCTTCCAG 3983
Qy 638 rThrAsnGlyGlySer-----SerTrpSerAlaIleThrGlyValSerSerAl 654
Db 3984 CGACGCGAACGCGATGTGTGACGTACGCTGGACGAGGTCTCCGCGCCGGAACGCGAC 4043
Qy 654 aValAsn 656
Db 4044 CATCTCC 4050
```

RESULT 14

```
US-09-902-540-1119
; Sequence 1119, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
```

```
; SEQ ID NO 1119
; LENGTH: 16584
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1119

Alignment Scores:
Pred. No.: 6,65e-05 Length: 16584
Score: 175.50 Matches: 165
Percent Similarity: 35.27% Conservative: 90
Best Local Similarity: 22.82% Mismatches: 240
Query Match: 4.35% Indels: 229
DB: 4 Gaps: 47

US-09-917-376-3 (1-740) x US-09-902-540-1119 (1-16584)

Qy 24 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg 43
Db 6448 TACAACCTCAGTTCCCGGCATGCTCCTCGACATCGACCTGGGCACGGCTACGCC 6507
Qy 44 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAsp-----TrpValGly 59
Db 6508 GCGAAGCGGTGACGTCCTCGGTTCGGGTGCGTACGAGAGCGCGGCTACACGGC 6567
Qy 60 Trp-----AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 77
Db 6568 TGGCTGTGAGGACGCTCGAGTTCAACGGCATC-----6600
Qy 78 ThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGly 97
Db 6601 ACGAACACTCGGTCGCCACCATCGC-----CCCGAGGATGGC 6639
Qy 98 AlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLys 117
Db 6640 GTTGT-----GTAAACCCCTGCGCCATCGC 6666
Qy 118 LeuGlyGlyAsnMetProGlyArgGlyMet-----GlyGluArgLeu-----131
Db 6667 AACGCGGT-----CCGACCGGTCCATCGCGCGGGTGAAGTGTGAGCCCTACGGT 6720
Qy 132 ---AlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 150
Db 6721 TCGGCTCGCATCGGAAGCGCGCGGTGACCTTC-----6756
Qy 151 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 170
Db 6757 -----ACCTGGGCACAG---ACGTCTGCGCTCGCGTGC 6786
Qy 171 -----GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIle 187
Db 6787 ACCCTGGCGGGTGCACACCTGAACCGCTGTTACCGCGCGCGCGGCTTACCGAGTCC 6846
Qy 188 GlnGlyValTrpValAlaPheAspLysSerSerSerSerSerSerSerSerSerSer 207
Db 6847 ACCGCACTGGTGTTCACGCTG-----ACCGTCTCGACGCGCTCAAGAGCTCCACGCG 6900
Qy 208 ThrIlePheValGlyValAlaAspProAsnProValPheTrpSerArgAspGlyGly 227
Db 6901 TCGGTGACCGGTGACGCTGGCGCTCCCAACATCCG-----6936
Qy 228 AlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisGlyValPhe 247
Db 6937 -----CCACCGTGAACCGGGCTCGACGGCATCTCGAGAGCGCGCTGAGTAC 6987
Qy 248 AspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGly 267
Db 6988 -----ACGCTGAGCGGCTCCGCCAGCGATGCGATGGC 7020
Qy 268 SerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIleSerPro 287
Db 7021 AATGCC-----CTCACGTACTGTGGAGCCCAAGGTCTCCGGT 7056
Qy 288 ValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArg 307
```


Db	7057	ACGCCGGTG-----			CGCGTGAAGGACTAC-----	7080
Qy	308	GlnHisProAsnThrIleMetValalaThrGlnIleSerTrpProAspThrIleIle	327			
Db	7081	ACCACCGCCAGCGGACGTTCACTCGCGCCGAGGTACAG-----CTGGATGATCGGTGGTTC	7137			
Qy	328	PheArgSerThr-----AspGlyGlyAlaThrTrp-ThrArgIleTrpAspTrpThrSe	345			
Db	7138	TTCCGGTCTACGGTTCAGCAGCGCATCGCCACGGTGAACG-----ACACG	7182			
Qy	345	rTrpProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPh	365			
Db	7183	GTCAACGGTCACGGTCACCAACGGA-----ACCGCGCGCCCA-----	7219			
Qy	365	eGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMe	385			
Db	7220	-----TCGTCTCCGACACGCTCCGTGGCGCTTCGCGCGGGCACT-GTCACCGGTACGGCGCTTC	7274			
Qy	385	tAlaIleAspProPheAsnSerAspArgMetLeuTyrGly-----ThrGlyAlaTh	402			
Db	7275	CGCGGTGACCCG---GATGGTGACCGCTCAGCTACAGCTGGGAGCAGACGGGTGGTTC	7331			
Qy	402	rLeuTyrAlaThrAsn-----	407			
Db	7332	GACGGTCGCCATCAATGGCGCGGACACGTCGCGCATCTCTTCGCCACCCCGTCCCGG	7391			
Qy	408	-----	408	-----	-----As	
Db	7392	TTCGTACGAGTTTCAACCGTGACGGCGGACGGCGCTCGGCTCTGCTCTCAAGCGGTGCC	7451			
Qy	408	pLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLys---GlyLe	427			
Db	7452	GGTGACCATCATCGATGGTTCTTCGCGGCGAACTCTGCGCCAGCCTGACCGCTGAAACCGGGCAT	7511			
Qy	427	uGluGluThrAla---ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSe	446			
Db	7512	CGATGCCACGGCGAAGCGTGTGACGTCTGACGTCTGACGCTCAGCGGCTCGCGGAGC-----GA	7565			
Qy	446	rAlaLeuGlyAsp-----	454		LeuGlyGly-----Ph	
Db	7566	CGCGAGGGGTGACACGCTGACGTACCACTCGGAGCAGATCGGTGGCAGCGATGTGACCCCT	7625			
Qy	454	eThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrTh	474			
Db	7626	GACGGGCGCCGAC---ACCTGTAGCGCGTCTG-----TTCAACCGCCGAGCAGCGCCAG	7676			
Qy	474	gIlyThrSerValAspTyrAlaGluLeuAsnProSer-----	486			
Db	7677	CGGTACCAACGCTCGGCTTCATC---CTGACGGTCAGCGATGGCACCTCGACCAACCAGCGA	7733			
Qy	487	-----IleIleValArgAlaGlySerPheAppProSerSerGlnProAsnAspAr	503			
Db	7734	TGCGGTGCGGCTCATCTGTGGCGG-----GACCCGGGTGCCAACCCGAGC-----	7779			
Qy	503	gHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyG	523			
Db	7780	-AACACTGCTCCGGAACGGACGGCGGCGAGTTCGCCATCGTG---CGGAGGGCGGCCAC	7835			
Qy	523	yValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg-----PheValTr	541			
Db	7836	TGTGACGCTCAACGGCAGC---GCGACCGACGCTGACGGTGACACGCTCGTCATCGTCTG	7892			
Qy	541	pAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAl	561			
Db	7893	GACGACG---ATCGGCGGACGCGCGGTG-----ACGCTCTGTGACGCGCTCTCGCTGAC	7943			
Qy	561	aAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysTh	581			
Db	7944	GCGAGCGTTCAACCGCCCGGCGCTCTCTC-----	7982			
Qy	581	rPheTyrAlaLeu---SerAsnGly-----ThrPheTyrArgSerThrAspGlyGlyVa	598			

```

7983  CTTCTCTCTGATGGTCAGCGACGGTACGGCGACTGTCGTGCGACGTGACGTGACGTCCATCAGCGT 8043
Qy      598  lThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHi 618
Db      8043  GACCGAGGAAAGCTCGCG-----CCGGTGGCGTCCGCTCCGGCGGTGCTC----- 8088
Qy      618  sAlaValProGlyGlyGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSe 638
Db      8089  -----TCTGGCAACCAAGACGTCGGCCACGCTCGACGGCTCGCGCTTCACG 8132
Qy      638  rThrAsnGlyGlySer-----SerTrpSerAlaAlaThrGlyValSerSerAl 654
Db      8133  CGACGGCAACGGCGATGTGTGACGTACCGTCCGACGCGAGGTCTCGGCGCCGACCGGAC 8192
Qy      654  aValAsn 656
Db      8193  CACTCC 8199

RESULT 15
US-09-453-702B-39
; Sequence 39, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
;            Burland, Valerie
;            Perna, Nicole T.
;            Plunkett, Guy
;            Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25165
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:

US-09-453-702B-39

Alignment Scores:
Pred. No.: 0.000375 Length: 25165
Score: 170.50 Matches: 192
Percent Similarity: 29.22% Conservative: 83
Best Local Similarity: 20.40% Mismatches: 269
Query Match: 4.22% Indels: 397
DB: 3 Gaps: 48

```

Alignment Scores:	
Pred. No.:	0.000375
Score:	170.50
Score Similarity:	29.22%
Best Local Similarity:	20.40%
Query Match:	4.22%
Matches:	1
Mismatches:	3
Indels:	1
Gaps:	1
Length:	25165
Matches:	192
Conservative:	83
Mismatches:	269
Indels:	397
Gaps:	48


```

QY 578 -----AsnProLysThrPheTyrAlaLeuSer-----AsnGly 588
Db 20332 CACACCTATAACCTGAACAACATATTACACGCTGCTTCGCTCATCAACCAGGGGAATGGG 20391
QY 589 ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 608
Db 20392 ACGTTTGTCTGGGGGAGACACCAACCAATACCTTCTCGAGCGGGCGGCAGTGGCGCT 20451
QY 609 SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys----GluGlyAspLeu 627
Db 20452 ATGAGCAGCAGCGTTTCCATGACCTGGCGCGATTCGATGGCGGCGATATGGATCTC 20511
QY 628 TrpLeuAlaAlaSerSerGly-----LeuTyrHisSerThrAsnGly 641
Db 20512 TTCTTGGCCGCCAGCCAGGGAAGAGCTAACTACGGCTCGCTGTTATTCAACACCAACCGC 20571
QY 641 ----- 641
Db 20572 GTACTGGGTTGCCCGGTGGCGGCGCAACGCCAACCACTACGCCAGCCAGTTTAGC 20631
QY 642 -----Gly 642
Db 20632 CTGGCGGTGGACTGGAACCAACGCGCTGATGGATATGCCCGTATCGCCAGACCGGG 20691
QY 643 SerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662
Db 20692 CAGTCGTAT---CTTTATATAACGTGACCAACGCCAGCAAC-----TGGACGCAATCG 20742
QY 663 AlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr--- 681
Db 20743 GCCCTCGCGCGCAGCCAGC-----GGTACCACCGAGCGGTGGCGCA 20787
QY 682 -----GlyValTyrArg-----SerAspAspCysGlyThrThr 692
Db 20788 ATGGACTACGACTGGGAGCGCGGTGGTGGTGTCCAAACAGTCGGGCGAGCGTG 20847
QY 693 TrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGly 712
Db 20848 TTCTGTAGCCGCAACACCAACACCGGTGAGCTACGGC-----ACTTCG 20889
QY 713 AspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGly 732
Db 20890 CTACACCTGGCGATC-----ACCGATCCCAACGGCATTAACGTCTATTACGGC 20937
QY 733 Asp 733
Db 20938 AAT 20940

```

Search completed: October 5, 2005, 22:56:47
Job time : 5746.23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 08:30:02 ; Search time 1425.51 Seconds
(without alignments)
3584.301 Million cell updates/sec

Title: US-09-917-376-3
Perfect score: 4036
Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGTNGRGIVYDGGAPSG 740

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09917376/runat_04102005_164329_26954/app_query.fasta_1.2524
-DB=Published Applications NA -QWTF=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USFR=US09917376 @CGN 1 1 2178 @runat_04102005_164329_26954
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Published Applications NA:
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
 - 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
 - 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
 - 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
 - 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
 - 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
 - 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
 - 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
 - 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
 - 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
 - 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
 - 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
 - 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
 - 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
 - 19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:
 - 20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:
 - 21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:
 - 22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
 - 23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:
 - 24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
 - 25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
 - 26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4036	100.0	2869	11	US-09-917-376-2	Sequence 2, Appli
2	4036	100.0	2869	15	US-10-155-400-2	Sequence 2, Appli
3	2429.5	60.2	2846	15	US-10-156-761-1845	Sequence 1845, Ap
4	2429.5	60.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
5	1579	39.1	2517	15	US-10-026-994-4	Sequence 4, Appli
6	1579	39.1	2710	15	US-10-026-994-1	Sequence 1, Appli
7	1405	34.8	2217	15	US-10-156-761-2561	Sequence 2561, Ap
8	1405	34.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
9	1139	28.2	3668	10	US-09-927-827-21	Sequence 21, Appli
10	1117	27.7	2367	18	US-10-395-241-13	Sequence 13, Appli
11	1117	27.7	2646	18	US-10-395-241-11	Sequence 11, Appli
12	1114	27.6	2481	18	US-10-395-241-17	Sequence 17, Appli
13	878.5	21.8	5698	18	US-10-420-191-1	Sequence 1, Appli
14	392.5	9.7	1103	20	US-10-653-047-7511	Sequence 7511, Ap
15	238	5.9	7407	16	US-10-246-330-3	Sequence 3, Appli
16	238	5.9	7407	17	US-10-282-122A-30151	Sequence 30151, A
17	199.5	4.9	7305	17	US-10-282-122A-11269	Sequence 11269, A
18	195	4.8	5295	17	US-10-282-122A-11871	Sequence 11871, A
19	192	4.8	9903	17	US-10-282-122A-28185	Sequence 28185, A
20	188.5	4.7	4399	17	US-10-369-493-37642	Sequence 37642, A
21	184	4.6	3324	17	US-10-282-122A-26289	Sequence 26289, A
22	182	4.5	69350	18	US-10-041-018-379	Sequence 379, App
23	180	4.5	2355	17	US-10-369-493-37792	Sequence 37792, A
24	179.5	4.4	3222	17	US-10-282-122A-14432	Sequence 14432, A
25	176	4.4	32367	14	US-10-158-160A-14	Sequence 14, Appli
26	174	4.3	2855	24	US-11-097-143-40748	Sequence 40748, A
27	174	4.3	5221	24	US-11-097-143-26683	Sequence 26683, A
28	174	4.3	5302	24	US-11-097-143-40747	Sequence 40747, A
29	173	4.3	3708	24	US-11-097-143-26684	Sequence 26684, A
30	170.5	4.2	2176	24	US-11-097-143-14408	Sequence 14408, A
31	170.5	4.2	4252	24	US-11-097-143-14407	Sequence 14407, A
32	170.5	4.2	7191	17	US-10-282-122A-35048	Sequence 35048, A
33	170.5	4.2	25165	14	US-10-114-170-39	Sequence 39, Appli
34	170	4.2	4323	17	US-10-282-122A-8383	Sequence 8383, Ap
35	169.5	4.2	3360	15	US-10-245-802-11	Sequence 11, Appli
36	169	4.2	69350	18	US-10-041-018-379	Sequence 379, App
37	168	4.2	5451	15	US-10-101-510-597	Sequence 597, App
38	167.5	4.2	2110	13	US-10-027-632-97984	Sequence 97984, A
39	167.5	4.2	2110	17	US-10-027-632-97984	Sequence 97984, A
40	167	4.1	1839	17	US-10-369-493-34868	Sequence 34868, A
41	166	4.1	3879	17	US-10-282-122A-8717	Sequence 8717, Ap
42	165.5	4.1	4393	15	US-10-245-802-21	Sequence 21, Appli
43	165	4.1	6615	17	US-10-282-122A-48180	Sequence 28180, A
44	165	4.1	67167	21	US-10-937-730A-3	Sequence 3, Appli
45	164.5	4.1	2540	22	US-10-680-860A-161	Sequence 161, App

ALIGNMENTS

RESULT 1
US-09-917-376-2
; Sequence 2, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1


```
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-2

Alignment Scores:
Pred. No.: 0 Length: 2869
Score: 4036.00 Matches: 740
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-09-917-376-3 (1-740) x US-09-917-376-2 (1-2869)

Qy 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20
Db 139 GCGACGACTCAGCGGTACACCTGGAGCAACGTGGCGATCGGGGGCGGCTTTGTTCGAC 198
Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40
Db 199 GGGATCGTCTCAATGAGGTGCACCGGGAATTCGTACGTGCGGACCGACATCGGGGG 258
Qy 41 MetTyrArgTTPAspAlaAlaAsnGlyArgTTPileProLeuLeuAspTTPValGlyTTP 60
Db 259 ATGTATCGATGGATGCGCCCAACGGCGGTGGATCCCTCTTCGTGGATTGGTGGGATGG 318
Qy 61 AsnAsnTTPGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 319 AACAAATGGGGGTACAAAGCGGTCTCAGCATTTGCGGACAGCCCGCATCAATTAACAAG 378
Qy 81 ValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAspProAsnAspGlyAlaIleLeu 100
Db 379 GTATGGGCGCGTGGGAATGTACCAACAGCTGGGACCCCAACGACGAGCGGATTC 438
Qy 101 ArgSerSerAspGlnGlyAlaThrTTPGlnIleThrProLeuProPheLysLeuGly 120
Db 439 CGCTCGTCTGATCAGGGGCAACGTGGCAATAACGCCCTCGCTTCAAGTTGGCGGC 498
Qy 121 AsnMetProGlyArgGlyMetGlyuArgLeuAlaValAspProAsnAsnAspAsnIle 140
Db 499 AACATGCCGGCGCTGGAAATGGCGAGCGGTGGTGGTGGATGCCAAACAATGACAACTT 558
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTTPArgSerThrAspSerGlyAlaThr 160
Db 559 CTGTATTTCGGCGCCCGAGCGCAAGGGCTCTGGAAGACACAGATTCCGGGCGGACC 618
Qy 161 TTPSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 619 TGGTCCAGATGACGAACCTTCGGACAGTAGGACGTACATTGCAATCCCACTGACACG 678
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValTTPValAlaPheAspLysSerSer 200
Db 679 ACCGCTATCAGACGATATTCAAGCGCTCGTCTGGGTGGTTCGATCGATTCGTC 738
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
Db 739 TCGTTCGGCGAAGCGAGTAAGACCAATTTTGTGGCGGTGGCGGATCCCAATAATCCGGTC 798
Qy 221 PheTTPSerArgAspGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe 240
Db 799 TTCGGACGAGACGGCGGCGGACGTGGCGAGCGGTGCCGGTGGCGGACCGGCTTC 858
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
Db 859 ATCCGACACAGGGGCTTTTGACCCCGGTCAACACGTTCTATATTGCCACACGACAT 918
Qy 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTTPLysPheSerValThrSerGly 280
Db 919 ACGGTGGTGGTATACGGGAGCTCCGGGACGCTCGGAAATTTCTCGGTGACCTCCGG 978
Qy 281 ThrTTPThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
Db 979 ACATGGACGCGAATCAGCCCGGTACCTTCGACGACACCGCGCGTGGAGCGCGATC 1038
Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
Db 1039 AGCGGCTCACTATCGACGCGCAGCACCCGACACGATAATGGTGGCAACCCAGATATCG 1098
Qy 321 TTPTTPProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTTPThrArgIle 340
Db 1099 TGGTGGCGGACACCATTAATCTTCGAGACACCGACGCGGTGGAGCGCGATC 1158
Qy 341 TTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
Db 1159 TGGGATTGGACGAGTTATCCCAATCGAAGCTTGGCATATGTCTTGACATTTCCGGGGAG 1218
Qy 361 ProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTP 380
Db 1219 CCTTGGCTGACCTTCGGGTACAGCCGAATCTCTCCGTACCCAGTCCGAAGCTCGGCTGG 1278
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
Db 1279 ATGGATGAAGCGATGGCAATCGATCGTTCAACTCTGATCGGATGCTCTACGGAACAGGC 1338
Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIle 420
Db 1339 GCGAGCTTGTACGCAACAAATGATCTCAGAAATGGGACTCCGGCGGCGCAGATTCAATC 1398
Qy 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 440
Db 1399 GCGCGATGGTCAAGGATTTGGAGAGACGGCGGTAAACGATCTCATACGCCCGCGTCT 1458
Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460
Db 1459 GCGCGCCGCTCATCAGCGCTCTCGAGACCTTCGGCGGTTCACCCACGCGGAGTTACT 1518
Qy 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
Db 1519 GCGGTGCCATCGACGATCTTCACGTCAACGGTGTTCACGACCGGACACGCGTCACTAT 1578
Qy 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500
Db 1579 GCGGAATTTGAATCCGTTCGATCATCTTCGCGTGGAGTTTCGATCATCGACCCAAACCG 1638
Qy 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTTPPheGlnGlySerGlu 520
Db 1639 AACGACGACGCTCGCTTCGACAGACGGCGGCAAGACTGGTTCCAAAGCAGCGAA 1698
Qy 521 ProGlyGlyValThrThrGlyThrValAlaIleAspAlaAspGlySerArgPheVal 540
Db 1699 CTTGGCGGGGTGACGACGGCGGACCGCTCGCGCATCGCGGACGCGCTCTCGTTCGTC 1758
Qy 541 TTPAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTTP 560
Db 1759 TGGGCTCCGCGGATCCCGGTGAGCTGTGGGTGACGCGAGTGGATTTGGCAATCTCTCG 1818
Qy 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580
Db 1819 GCTGTTTCGAAGGTGTTCCCGCAATGCCAGATCCGCTCAGACCGGGTGAATCCAAG 1878
Qy 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600
Db 1879 ACTTTCTATGCCCTATCCAAATGGAACCTTCTATCGAAGCAGCAGCGCGCGGTGACATTC 1938
Qy 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 620
Db 1939 CAACCGGTCCGCGCGGTCTTCGAGCAGCGGTGCGGTGCTCATGTGTTCCACGCGGTG 1998
Qy 621 ProGlyLysGluGlyAspLeuTTPLeuAlaIleSerSerGlyLeuTyrHisSerThrAsn 640
```


Db 1999 CCTGGAAAGAGCGCATCTGTGGCTCGCTGCATCGAGGGGCTTTACCACTCAACCAAT 2058
 Qy 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660
 Db 2059 GCGCGACAGTGTGTCTGCAATACCGCGGTATCCTCCGGGTGAACGTGGGATTTGGT 2118
 Qy 661 LysSerAlaProGlySerSerTrpProAlaValPheValValGlyThrIleGlyVal 680
 Db 2119 AGTCTGCGCGCGGTGTGTATACCGACCGCTTTGTCTGCGCACGATCGGAGCGTT 2178
 Qy 681 ThrGlyAlaTrpArgSerAspCysGlyThrTrpValLeuIleAsnAspGln 700
 Db 2179 ACGGGGCGTACCGCTCCGACGACTGTGGACGACCTGGGTACTGATCAATGATGACCG 2238
 Qy 701 HisGlnTrpGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720
 Db 2239 CACCAATACGGAATTTGGGGACAAGCAATACCGGTGACCGCAATTTACGGCGGGTG 2298
 Qy 721 TyrIleGlyThrAsnGlyArgGlyIleValTrpGlyAspIleGlyGlyAlaProSerGly 740
 Db 2299 TACATAGACGACGACCGCGTGGAAATTTGTATACGGGGACATTTGGTGGTGGCCGTC 2358

RESULT 2

US-10-155-400-2
 ; Sequence 2, Application US/10155400
 ; Publication No. US20030108988A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DING, SHI-YOU
 ; APPLICANT: ADNEY, WILLIAM S.
 ; APPLICANT: VINZANT, TODD B.
 ; APPLICANT: HIMMEL, MICHAEL E.
 ; TITLE OF INVENTION: THERMAL TOLERANT AMYLASE FROM ACIDOTHERMUS
 ; FILE REFERENCE: CELLULOYTICUS
 ; CURRENT APPLICATION NUMBER: US/10/155,400
 ; CURRENT FILING DATE: 2002-10-22
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 2869
 ; TYPE: DNA
 ; ORGANISM: Acidothermus cellulolyticus
 ; FEATURE:
 ; NAME/KEY: modified_base
 ; LOCATION: (2869)
 ; OTHER INFORMATION: a, c, t, g, other or unknown
 US-10-155-400-2

Alignment Scores:
 Pred. No.: 0 Length: 2869
 Score: 4036.00 Matches: 740
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 15 Gaps: 0

US-09-917-376-3 (1-740) x US-10-155-400-2 (1-2869)

Qy 1 AlaThrThrGlnProTrpTrpSerAsnValAlaIleGlyGlyPheValAsp 20
 Db 139 GCGACGACTCAGCCGTACACCTGGAGCAACGTGGCGATCGGGGCGCGGCTTTGTCGAC 198
 Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTrpValArgThrAspIleGlyGly 40
 Db 199 GGGATCGCTTTCATGTAGAGGTGACCGGGAAATCTGTAGTGGGACGACATCGGGGG 258
 Qy 41 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60
 Db 259 ATGTATCGATGGATGCCCAACGGGCGTGGATCCCTCTTCTGGATTGGGTGGGATGG 318
 Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
 Db 319 AACAAATGGGGGTACAAACGGCGTGTGATTTGGAGGAGACGGCGTAAACGATCTCATCAGCCCGCGCT 1458

Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
 Db 379 GTATGGGCGCGCTCGGAATGTACCAACAGCTGGACCCAAACAGCAGCGAGCGATTCTC 438
 Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120
 Db 439 CGCTGCTGTGATCAGGGCGCAACGTGGCAAAATACCCCTTCGCTTCAAGCTTGGCGGC 498
 Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
 Db 499 AACATGCCCGCGCGTGGAAATGGCGAGCGCTTGGGTGGATCCAAACAATGACCAATT 558
 Qy 141 LeuTrpPheGlyValProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
 Db 559 CTGTATTTCCGGCGCCCGAGCGCAAGGGCTCTGGAGAAGCACACAGATTCCGCGCGACC 618
 Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTrpIleAlaAsnProThrAspThr 180
 Db 619 TGGTCCCAGATGACGAACCTTCCGGACGTAGGCACGTACATTGCAAAATCCCACTGACAG 678
 Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200
 Db 679 ACCGGCTATCAGAGCGATATTCAAGCGCTCGTCTGGGTGCTTTCGCAAGTCTTCGTCA 738
 Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
 Db 739 TCGCTCGGCAAGCGAGTAGACCATTTTGTGGGCTGGCGGATCCCAATAATCCGGTTC 798
 Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240
 Db 799 TTCTCGAGCAGAGACGGCGCGACGTGGCAGCGGTGCGGGTGGCGGACCGCGCTTC 858
 Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTrpIleAlaThrSerAsn 260
 Db 859 ATCCCGCAACAGGGCGCTTTGACCCGCTCAACACGCTGCTTATTTGTCACCGACCAAT 918
 Qy 261 ThrGlyGlyProTrpAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280
 Db 919 ACGGGTGGTCCGTATGACGGGAGCTCCGGCGAGCTCTGGAAATTTCTCGGTGACCTCGG 978
 Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTrpPheGlyTyr 300
 Db 979 ACATGGACCGGAATCAGCCCGGTACCTTCGACGACACGCGCGGCTACTTTGGTTAC 1038
 Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
 Db 1039 AGCGGCTCACTATTCAGCGCCACGACCCGAAACACGATATATGGTGGCAACCCAGATATCG 1098
 Qy 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340
 Db 1099 TGGTGGCGGACACCATTAATCTTTTCGGAGCACCGACGCGGTGGAGCTGGACCGGATC 1158
 Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
 Db 1159 TGGGATTGGACGAGTTATCCCAATCGAAGCTTGGGATATGCTTGACATTTGGCGGAG 1218
 Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrp 380
 Db 1219 CCTTGGCTGACCTTCCGGGTACAGCGCAATCTCTCCGTTACCCAGTCCGAAAGCTCGGTGG 1278
 Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTrpGlyThrGly 400
 Db 1279 ATGGATGAAGCGATGCAATTCGATCCGTTCAACTCTGTATCGAATGCTTACGGAACAGC 1338
 Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
 Db 1339 GGGACCTTGTACCCAAACAATGATCTCACGAGTGGGACTCCGCGCGCCAGATTCAATC 1398
 Qy 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 440
 Db 1399 GCGCCGATGGTCAAGAGGATTGGAGGAGACGGCGTAAACGATCTCATCAGCCCGCGCTCT 1458

441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 460
Db GCGCCCGCTCATCGGCTCTCGAGACCTCGGGCTTCAACCGCGAGCTTACT 1518
461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480
Db GCGGTGCATCGAGCATCTTCACGTCACTGCTTTCAGACCGGCGAGCGTGCATAT 1578
481 AlaGluLeuAenProSerIleValArgAlaGlySerPheAspProSerSerGlnPro 500
Db GCGGAATGAATCCGTCGATCATGCTTCGCTCGAAGTTTCGATCCATCCAGCCACCG 1638
501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu 520
Db AACGACGCGACGTCGCTTCGACACGCGGCGAAGACTGTTTCCAAAGGCGAGAA 1698
521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540
Db CCGCGCGGTGACGACGCGGCGCACCGCTCGCGCATCGGCGAGCGGCTCTCGTTTCGTC 1758
541 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 560
Db TGGGCTCCGCGCATCCCGGTGAGCTGTGTGTACGCGAGTCCGATTTGGCAACTCTCTGG 1818
561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580
Db GCTGCTTCGCAAGGTGTTCCGCCAATGCCAGATCCGCTCAGACCGGCTGAATCCAAAG 1878
581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600
Db ACTTCTATGCTTATCCAAATGAACCTTCTATCGAAGCAGCGCGCGGTGCATTC 1938
601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValAlaGlyValMetPheHisAlaVal 620
Db CAACCGGTGCGCGCGCGGTCTTCGAGCAGCGGTGCGGTGCTCATGTTCCACGCGGTG 1998
621 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 640
Db CTGGAAGAAGAGCGCATCTGTGGCTCGCTGATCGACGCGGCTTTACCATCAACCAAT 2058
641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660
Db GCGGCGACGATGCTGCTGCAATCACCGCGGTATCTCCGCGGTGAACGTGGATTTGGT 2118
661 LysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGlyVal 680
Db AAGCTGCGCGCGGTGCTGATACCGACCGCTTCTTGTGCGGCGACGATCGGAGCGGT 2178
681 ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGln 700
Db ACGGGGCGGTACCGCTCCGACGACTGTGGAGCAGCTGGGTACTGATCAATGATGACCCAG 2238
701 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgVal 720
Db CACCAATACGGAAATTTGGGCAACAGCAATCACCGGTGACCGCAATTTACGGCGGTG 2298
721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740
Db TACATAGGCAGCAACGCGCGTGGAAATTGTATACGGGAGCAATTTGTGTGTCGCGCGT 2358

RESULT 3
US-10-156-761-1845
; Sequence 1845, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1845
LENGTH: 2646
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2646)
US-10-156-761-1845
Alignment Scores:
Pred. No.: 4,45e-224 Length: 2646
Score: 2429.50 Matches: 438
Percent Similarity: 73.9% Conservative: 111
Best Local Similarity: 59.0% Mismatches: 174
Query Match: 60.2% Indels: 19
DB: 15 Gaps: 8
US-09-917-376-3 (1-740) x US-10-156-761-1845 (1-2646)
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
Db 115 TACAGTGGAGAAAGCCCGCGTGCAGCGCGCGGCTTCGTCCCGGCATCGTCTTCAAC 174
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45
Db 175 CGTCCGAGAGAACTCGCTACCGCCGACCGATCGCGCGGCTTACCGCTGGGCC 234
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
Db 235 GAGTCTCGAAGACCTGGAGCGGCTGTCTCGACTCGTCTGGTGGAGCGACTGGGGGCAC 294
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAlaAlaVal 85
Db 295 ACGGTGTGTGAGGCTCGCTCCGACTCCGCGACCGCCGAAAGAGGTGTACGCGGCGCTC 354
QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
Db 355 GGCAGTACAGAAACAGCTGGGACCGGGCAACGCTGCTGTCTAGTCCGCGCGACCGG 414
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
Db 415 GCGCGAGCTGGCAGAGACCGACCTGCCCTTCAAGCTGGCGGGAAACATCGCGGCGCG 474
QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
Db 475 GGCATGGGCGAGCGGCTCGCGTGCACCGGAAACAGGACAGCGGTGTGTATTCGCGCGG 534
QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
Db 535 CCCACGGCAAGGGCTGTGGCGTGCAGCGACTCGGGGCGCTTCTGGTTCGACAGTCAACC 594
QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185
Db 595 GACTTCCCGAAGCTCGGCACCTACCTGCGAGGACCGGACCGACAGCGGGTACCGGTCC 654
QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205
Db 655 GACAACACAGGGCATCGTGTGGGTCACTTCGACGAGTTCGACGCGGGTTCGCGGGGAGTCC 714
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValThrTrpSerArgAsp 225
Db 715 ACGCGAGCGGTGTACGTGCGGGTCCCGGACAAAGCAACTCCGTCTCTATCGCTCCACGGAC 774
QY 226 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly 245

775 GCGGCGGACCTGGTCCCGGCTGGCGGCGGACCGGCGCATCTCGCCCAAGGCG 834
246 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 265
835 GTGCTGGACGCGGCGAAGCGGCTGTCTGTACCTCGGTACAGGACGAGGCGCGACCGTAC 894
266 AspGlySerSerGlyAspValTyrIlePheSerValThrSerGlyThrThrThrArgIle 285
895 GACGCGGCGGAGGACGCTGTGGGCTGACAGCAGGACCGGACCGTGGACGACATC 954
286 SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 305
955 AGCCCGGCTCGGAGCGGACAC-----TACTACGGCTTCAGCGGGCTGACCGGTG 1005
306 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerThrThrProAspThr 325
1006 GACGCGGACATCCGCGGAGCGGTGTGGCGACTGCGTACAGCTTCCTGGTGGCGGACG 1065
326 IleIlePheAspSerThrAspGlyAlaThrThrArgIleTyrAspThrThrSer 345
1066 CAGCTTCTCCGCTCCAGGACAGCGCGGACCTGGACGAGGCTGGGACTACACTCG 1125
346 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTyrLeuThrPhe 365
1126 TATCCGAGCGCTCGAACCGCTTCACCATGGATGTCTCTCGCTCGGCTGGCTCACCTGG 1185
366 GlyValGlnProAsnProValProSerProTyrLeuGlyTyrMetAspGluAlaMet 385
1186 GGAGGCGACCGCGCGGCGGACAGCCCGGAACTCGGCTGGATGACCGAGTCCCTG 1245
386 AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405
1246 GAGATCGACCGCTTCGCGCGCATGATGATCGGACCGGCGCGGCTTCACGCG 1305
406 ThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425
1306 ACGGACACTGACGACTGGGACGACGAGGAGCAGTTCACCATCAAGCGGATGGCGCG 1365
426 GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu 444
1366 GGCTGTGAGGAGCGGCGCTCAACGACCTCGCTCGCTCGCTCGGCGCGCGCCAGCTG 1425
445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464
1426 TTCAGCGCTCGGTGACATCGCGGCTTCGCGCACACGACGACCTCACCGGTGGCGCTG 1485
465 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 484
1486 CTGATGTACAGTTCGCGGAACTTCACACGACGACGACGACCTCGACTACGCGGACGCG 1545
485 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 504
1546 CCGGCGACGCTGGTGGGCTGGCGCACTTCGAC-----TCGGGTGCG-----CAT 1590
505 ValAlaPheSerThrAspGlyGlyLysAsnThrPheGlnGlySerGluProGlyVal 524
1591 GTGGCGTTCGACGACGACGCGGCGCACTGGTTTCGCGGGGCGGACCGCTTCGCGGGTC 1650
525 ThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyrAlaProGly 544
1651 AGCGGGGGTGGGACGCTGGCGGCGGTTCGACGCGGAGTCTGCTGTGGAGACCGCGCG 1710
545 AspProGlnProValValTyrAlaValIleGlyPheGlyAsnSerThrTrpAlaAlaSerGln 564
1711 GGACCGCG-----GTGCGATACACGACCGGTTTCGACCTCTGCTGGTGGCTCGCG 1764
565 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 584
1765 GGCTCTCGGCGGCGGATCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1824
585 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 604
1825 TTCAAGTCCGCGAGGTCTTACGTGATTCGACGCGGCGGCGGACCTTCACGCGGCTCGCG 1884

605 Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 623
1885 GCCACGGGCTCCGAGCGGCGGACAGC-----GTGCGCTTCAAGGCGCTGCCGCGCAG 1938
624 GluGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHisSer 638
1939 AAGGCGGACATCTGGCTGGCGGCGGCGAGCGAGCGGCGGTACCGGGCTGTGGCACTCG 1998
639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658
1999 ACGGACGGGCGGCGGCTTCACCAAGCTCGCCACCGTCCGACCGGCGGACACCATCGGC 2058
659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly 678
2059 TTGCGCAAGCGGCGGACCGGCGCTCTACACAGCGCTCTACACAGCGGCGGAGATCGGC 2118
679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp 698
2119 GGTGTGGCGGATCTTCGCGTCCGACGACGAGCGGCGGAGCTGGACCGCGCTCAACGAC 2178
699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718
2179 GATGCCACCGAGTGGGCTTGGACGGCGGCGGATCACCGGTGACCGGCTTACGGG 2238
719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro 738
2239 CGGCTGTGTGTGTCGACGAGCGGCGGCGGATCGTCTACGCGGACCGCGGCTCTTCG 2298
739 SerGly 740
2299 GACGGC 2304
RESULT 4
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Score: 1, 41e-219 Length: 9025608
Percent Similarity: 2429.50 Matches: 438
Best Local Similarity: 73.99% Conservative: 111
Query Match: 59.03% Mismatches: 174
DB: 60.20% Indels: 19
Gaps: 8
US-09-917-376-3 (1-740) x US-10-156-761-1 (1-9025608)

QY	6	TyrThrTrpSerAsnValalaileGlyGlyGlyGlyPheValaspGlyIleValPheAsn	25
Db	2277508	TACAGCTGAAGAAGCCCGGCTGCAGCGGCGGCTTCGTCCCGGATCGTCTTCAAC	2277567
QY	26	GlulGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp	45
Db	2277568	CGCTCCGAGAGAAGACCTCGCTACGCCCGCGCACCGACATCGCGGGCGCTACCGCTGGGCG	2277627
QY	46	AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr	65
Db	2277628	GAGTCTCGAAGACCTGAGCGCGCTGCTCGACTCGGTGCGGTGAGCGACTGGGGGCAC	2277687
QY	66	AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal	85
Db	2277688	ACGGGTGTCTGAGGCTCGCTCCGACTCCGACTCCGACCCCGAACAAGGTGTACGCGGCCGTC	2277747
QY	86	GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln	105
Db	2277748	GGCAGGTACAGAAACAGCTGGGACCCGGGCAACGGTGCCTGCTCAGGTCCGGCGACCGG	2277807
QY	106	GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg	125
Db	2277808	GGCGGAGCTGGCAGAAGACCGACCTGCGCTTCAAGCTGGGCGGGAACATGCGCGGCGCG	2277867
QY	126	GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla	145
Db	2277868	GGCATGGGCGACCGCTCGCGGTGCAGCCCGAACAAGACAGAGCTGCTGTATCTCGCGCG	2277927
QY	146	ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr	165
Db	2277928	CCACGCGCAAGGGCTGTGGCGGTGCAGGACTCGGGGCGCTCTGTGTCGAGGTCAAC	2277987
QY	166	AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGlnSer	185
Db	2277988	GACTTCCCGAAGCTCGCGCACTACGTGCAGGACCGGACCGACACAGCGGGTACGCGTCC	2278047
QY	186	AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla	205
Db	2278048	GACAACCGAGGATCTGTGGGTGCACCTTCGACGAGTGCACGGGTGCGCGGGAGCTCC	2278107
QY	206	SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp	225
Db	2278108	ACGCGGACGGTGTACGTGGGGTGGCGGACAGGACNACTCGTCTATCGCTCCACGAC	2278167
QY	226	GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly	245
Db	2278168	GCGGCGCGACCTGTGCTCCGGCTGGCGCGCGCACCGGCATCTCGCCCAACAAGGCG	2278227
QY	246	ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr	265
Db	2278228	GTGCTGGACCGCGGACCGGTGTGTGTACCTCGCGGTACAGCGACAAGGGCGGACCGTAC	2278287
QY	266	AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle	285
Db	2278288	GACGCGGCAAGGGACAGCTGTGGCGGTACACGACGAAGACCGGGACCTGGAGCAATC	2278347
QY	286	SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle	305
Db	2278348	AGCCCGGTGCGGAGCGCGACAC-----TACTACGGGTCTAGCGGGCTGACCGTG	2278398
QY	306	AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr	325
Db	2278399	GACCGGACGATCCGGGACCGGTGTGTGTACCTCGCGGTACAGCTCCTGGTGGCGGACACG	2278458
QY	326	IleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspTrpThrSer	345
Db	2278459	CAGCTCTTCCTTCCAGGACAGCGCGGACCTGGACGAAGGCCTGGGACACTACACTCG	2278518
QY	346	TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPhe	365
Db	2278519	TATCCGAGCGCTCGAAGCCCTTCAACATGGATGTCTCGTCTCGCCCTGGCTCACTCG	2278578
QY	366	GlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAlaMet	385
Db	2278579	GGAGGAACCCCGCACCCCGGAGCAGACCCGAAACTCGGTGGATGACCGAGTCCCTG	2278638
QY	386	AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla	405
Db	2278639	GAGATCGACCCGCTCGACTCCCGCGCATGATGATCGGAACGGGCGCGCTCTACGGC	2278698
QY	406	ThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys	425
Db	2278699	ACGGACAACCTGACCACTGGGACAGCGAAGCCACTTCAACATCAAGCCGATGGCGCG	2278758
QY	426	GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu	444
Db	2278759	GGCCTGGAGGAGACGGCGCTCAACGACTCGCTCGCTCCCTCCGCGCGCGCCAGCTG	2278818
QY	445	IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer	464
Db	2278819	TTACAGCGCGCTCGGTGACATCGCGCGCTTCCGCGCACACGGAACCTCAACCGGTGCGGTG	2278878
QY	465	ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn	484
Db	2278879	CTGATGTATACGCTCCCGGAACTTCAACACAGACACAGACTCGACTACGCCGAGACCGAC	2278938
QY	485	ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis	504
Db	2278939	CCGGCAGCGGTGTCGGGTCCGCAATCTCGAC-----TCGGGTCCG-----CAT	2278983
QY	505	ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal	524
Db	2278984	GTGGGTGTTCTCGACGACAAACCGCGCAACTGTGTTCGCGGGCGGACCCCTCGGGGCTC	2279043
QY	525	ThrThrGlyGlyThrValAlaAlaAspAlaAspGlySerArgPheValTrpAlaProGly	544
Db	2279044	ACCGGGGGTGGACCGGTCCGCGCGCTCCGCGCAGTCTCTGTGTGGAGCCCGCGCG	2279103
QY	545	AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln	564
Db	2279104	GGCACCGGG-----GTGCAGTACACGACCGGTTCGGCACCTCGTGGTCCGCGTCCGCG	2279157
QY	565	GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla	584
Db	2279158	GGCTCTCCCGCGCGGCGCATCGTCGAGTCCGACCGGGTCCGACCGAAGACTTCTTACGCGC	2279217
QY	585	LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla	604
Db	2279218	TTCAAGTCCGCGAGGTTCACGTTCAGTTCGACCGCGGGCGGACCTTACCGGCTCCGCG	2279277
QY	605	Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys	623
Db	2279278	GCCACGGGCTGCCGAGCGCGCGACAGC-----GTGCGCTTCAAGGCGCTGCCGCGCACG	2279331
QY	624	GlulGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHisSer	638
Db	2279332	AAGGGCGACATCTGGTGGCGGCGCGGAGCGAGCGCGCGGTGTGGCACTCG	2279391
QY	639	ThrAsnGlyLysSerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly	658
Db	2279392	ACGGACCGCGCGCGGCTTCACCAAGCTCCGACCGCGCGGCGGACCCATCCGCGC	2279451
QY	659	PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly	678
Db	2279452	TTCCGCAAGCGCGGACCGCGGCTCTGTACACAGCGCTCTACACCGCGGAGTCCGCG	2279511
QY	679	GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp	698
Db	2279512	GGTGTGCGGGCATCTTCGCTCGACCGCAAGGGCGCGAGTGGACCCCGCTCAACGAC	2279571
QY	699	AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg	718
Db	2279572	GATGCCACACCGGTGGAGCGGGCGGCGCATCACCGGTGACCCCGGCTTACCGG	2279631
QY	719	ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro	738

Db 2279632 CGCGTGTATGTGTCACGAACGGCGCGGATCTCTACGGCAGACCGCGGCTCTTCG 2279691
 QY 739 SerGly 740
 Db 2279692 GACGCG 2279697

RESULT 5

US-10-026-994-4
 ; Sequence 4, Application US/10026994
 ; Publication No. US20030113732A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dunn-Coleman, Nigel
 ; APPLICANT: Goedegebuur, Frits
 ; APPLICANT: Ward, Michael
 ; APPLICANT: Yao, Jian
 ; TITLE OF INVENTION: EGVI Endoglucanase and Nucleic Acids
 ; TITLE OF INVENTION: Encoding the Same
 ; FILE REFERENCE: GC698
 ; CURRENT APPLICATION NUMBER: US/10/026,994
 ; CURRENT FILING DATE: 2002-04-30
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 2517
 ; TYPE: DNA
 ; ORGANISM: Trichoderma reesei
 US-10-026-994-4

Alignment Scores:
 Pred. No.: 5, 7e-142 Length: 2517
 Score: 1579.00 Matches: 325
 Percent Similarity: 59.26% Conservative: 123
 Best Local Similarity: 42.99% Mismatches: 264
 Query Match: 39.12% Indels: 44
 DB: 15 Gaps: 20

US-09-917-376-3 (1-740) x US-10-026-994-4 (1-2517)

QY 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAspGlyIleValPhe 24
 Db 61 TTTTCATGGAAGACGTCAGCTCGCGCGCGGCTTCGTCGCCGCGCATCTTC 120
 QY 25 AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44
 Db 121 CATCCCAAGACAAAGGCGTAGCATATGCAGCAAGATATTGGCGGCTGTACCGCTC 180
 QY 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60
 Db 181 AAC---GCCGAGACTCATGGACCCCGCTCAGGATGGGATTCGTGATATCGCGGCTGG 237
 QY 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
 Db 238 CACACTCG-----GGCATCGACGCTGTGGCTGTGATCCGCGAGCATCAAAAG 288
 QY 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
 Db 289 GTGTATCGCGAGTCGGCATGTATACGAACAGCTGGGATCCGAGTAATCGAGCCATCAT 348
 QY 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120
 Db 349 CGCTGTCGACGCGCGGCGCAACGCTGCTTCACCACTTCCCTTCCTCAAGTCGGGGGT 408
 QY 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
 Db 409 AACATGCCAGGACGGGACCGGAGAGCGCTGTGGCTGTGATCCGGCCAACTCCAACATC 468
 QY 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
 Db 469 ATCTACTTTGGTGTCTCGCTCAGGAAACGGCTCTGGAGTCTACGGACGGCGGTGACC 528
 QY 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180

Db 529 TTTTCCAAAGTCTCGTGTTCACGGCAACTGGGACGATCATCCACACCGAGTGATCC 588
 QY 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200
 Db 589 AACGGCTACACAGCGCAGCAAGCAAGGACTCATGTGGTTACGTTCCGACTCAACGAGCAGC 648
 QY 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro----AsnAsnPro 219
 Db 649 ACGACCGGGGAGCCACGCTCTGTTGTTGGCAGCGCTCATAAACATCATCTGCTTCA 708
 QY 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239
 Db 709 GTCTATGTGACGACGAATCCCGCTCCAGTGTGAGTGTACCGGGGAGCGGAGAAA 768
 QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259
 Db 769 TACTTTCTCCACAAAGCGGAAACTGACGCCAGCAGAGAGGCTTGTATCTGACCTATTC 828
 QY 260 AsnThrGlyGlyProTyrAspGlySerGlyAspValTrpLysPheSerValThrSer 279
 Db 829 GATGGCAGCGGCGGTATGATGGCACACTTGGCTCAGTGTGGAGGTACGACATTCGAGGG 888
 QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
 Db 889 CGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAGATCTA-----TACTTTGGC 939
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
 Db 940 TTTGGCGCTTGGCCTCGATTTGCAAAAGCCAGGAACCCCTTGTGTGTCTTCTTTGAAC 999
 QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpTrpArg 339
 Db 1000 TCTTGTGGCCAGATGCTCAGCTGTTTCGGTCCGCCGCTCTGGGACACATCGAGCCG 1059
 QY 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359
 Db 1060 ATCTGGCGCTGGCGGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAAA 1119
 QY 360 GluProTrpLeuThrPheGly---ValGlnProAsnProValProSer----- 375
 Db 1120 GCACCGCTGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCGTCGATGGTCTC 1179
 QY 376 ---ProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394
 Db 1180 ATCAACGGCTCGCTCGGTGGATGATGATGCTCTCGAGATTGACCCACCGCAGCAGCAACC 1239
 QY 395 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 414
 Db 1240 TGGCTCTACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACAG 1299
 QY 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 434
 Db 1300 CGCCACAATGTGATCCAATCACTGGCAGAGCGGCATCGAGGAATTCCTCGTCCGAGAC 1359
 QY 435 LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 454
 Db 1360 CTGGCTCTGCACCCGGGAGCGAGCTATTGGCCGAGTCGGAGACGACGACGCTTC 1419
 QY 455 ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal 471
 Db 1420 ACCTTTGCCAGCAGAACGACCTCGGACATCCGCGCAGACGCTCTGGGCAACGCCACA 1479
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491
 Db 1480 TGGGCCACCTTCGACGAGCGTCTGACTACCGCGGAACTCGGTCAAGAGCGTCTCGCGCTC 1539
 QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
 Db 1540 GGCAACACCGCCCGCAGCAACAG-----GTGGCCATCTCTGTCGACGCGC 1584
 QY 512 GlyLysAsnTrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAla 531
 Db 1585 GGGCGCAGCTGGAGCATCGACTACCGCGGCCGACACGCTCCATGAACGGCGGCGACGCTGCC 1644


```
QY 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551
Db 1645 TATTTCGGCCGACGCGACGATCCTCTGTGTCGACCGCTCGTCGGG-----GTG 1695
QY 552 TyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln 571
Db 1696 CAGCGCTCGAGTTCACAGGCGAGTTTGCTCCGCTCGAGCCTCGCCGGCGCGCTC 1755
QY 572 IleArgSerAspArgValAsnProGlyThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591
Db 1756 ATCGCTCGGACAAAGACCAACAGCGCTCTTACGCGCGCTCGGATCGACCTTTTAC 1815
QY 592 ArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611
Db 1816 GTCAGCAAGGACACCGGACGAGCTTC-----ACGCGCGG-----CCCAAGCTGGGC 1863
QY 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629
Db 1864 AGCGAGGAGATCCGGGATATCGTCTCACCGACACCGCGGGACGTTGTATGTC 1923
QY 630 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyLysSerTrpSerAlaAla--- 648
Db 1924 TCGACCGAGCGCGCATATTCCGCTCCACAGACTCGGCGACGACCTTTGGCCAAAGTCTCC 1983
QY 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyr 668
Db 1984 ACCCGCTGACCAACACTACAGATCGCCCTGGGTGTGGGCTCA---GGCTCGAAGCTGG 2040
QY 669 ProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla-----TyrArgSer 686
Db 2041 ---AACCTGTATGCTTCGGACCC-----GGCCGCTCAGGGCTCGGCTCTAGCCAGT 2091
QY 687 AspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 706
Db 2092 GGAGACAGCGCGCTCTCTGACGACATCAGGCGCTCCAGGGCTCGGCTCCATCGAC 2151
QY 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGly 726
Db 2152 AGCACCAAGTTCGCGCGCGCGACGACACCGCCCGGCAAGTCTACGTGGGCGACCAACGGC 2211
QY 727 ArgGlyIleValTyr-----GlyAspIleGlyGlyAlaProSerGly 740
Db 2212 CGGGCGCTTTTACGTCAGGGAACCGTCGCGCGCGCGCGCGG 2259

RESULT 6
US-10-026-994-1
; Sequence 1, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGV1 Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-994-1

Alignment Scores:
Pred. No.: 6,27e-142 Length: 2710
Score: 1579.00 Matches: 325
Percent Similarity: 59.26% Conservative: 123
Best Local Similarity: 42.99% Mismatches: 264
Query Match: 39.12% Indels: 44
```


Db 1100 ATCTGGCGGTGGCGAGCTATCCGACTGAGACCTATTACTACGACATCTCAACTCCCAA 1159
QY 360 GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSer----- 375
Db 1160 GCACCGTGGATCAGAACAACATTTATCGATGTGACGAGCATCCGCTCGATGGTCTC 1219
QY 376 ---ProLysLeuGlyTrpMetAspGluAlaMetAlaLeuAspProPheAsnSerAspArg 394
Db 1220 ATCAAGCGCCTCGGTGGATGATTGAGTCTCTCGAGATTGACCCCAACGACGACCAACAC 1279
QY 395 MetLeuTrpGlyThrGlyAlaThrLeuTrpAlaThrAsnAspLeuThrLysTrpAspSer 414
Db 1280 TGGCTCTACGACCGCGAATGACATCTTGGCGGCCACGATCTCAACCACTGGGACACG 1339
QY 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434
Db 1340 CGCCACAATGTCTCAATCAATCACTGCGCAGACGCGCATCGAGGAATCTCGCTCCAGGAC 1399
QY 435 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPhe 454
Db 1400 CTGGCCTCGACCGCGGGAAGCAGCTATTGGCGCGAGTCTCGAGACGACCAACCGCTTC 1459
QY 455 ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal 471
Db 1460 ACCTTGGCAGCAGAAACGACCTCGGCACATCGCGCAGACGCTCTGGCAACGCCACA 1519
QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleValArgAla 491
Db 1520 TGGGCCACCTCGACGAGCGTGACTACGCGGGAACCTCGGTCAAGAGCGTCTCGCGCTC 1579
QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
Db 1580 GGCAACACCGCGCGCACCAACAG-----GTGGCCATCTCGTCCGACGCGC 1624
QY 512 GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAla 531
Db 1625 GCGCGAGTGGAGCATGACTACGCGCGCGACAGCTCTGTGTCGACCGCTCTCGCGC 1684
QY 532 AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal 551
Db 1685 TATTGGCGCGAGCGGACACGATCTCTGTGTCGACCGCTCTCGCGC-----GTG 1735
QY 552 TyrAlaValGlyPheGlyAsnSerTrpAlaSerGlnGlyValProAlaAsnAlaGln 571
Db 1736 CAGCGCTCGCAGTTCCAGGGCAGCTTGTCTCGTCTCGAGCTCGCGCGCGCGCGCTC 1795
QY 572 IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 591
Db 1796 ATCGCTCGGACAGAAGACCAACAGCGCTCTTCTACGCGCGCTCGGATCGACCTTTTAC 1855
QY 592 ArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 611
Db 1856 GTCAGCAGGACACCGGACGAGCTTC-----ACGCGGGG---CCCAAGCTGGGC 1903
QY 612 AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629
Db 1904 AGCGCAGGACGATCCGGATATCGTGTCTACCCGACCCAGCGCGGACGCTGTGTATGTC 1963
QY 630 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 648
Db 1964 TCGACCGAGCTCGGATATTTCCGCTTCCACAGACTCGGCGACGACCTTTGGCCCAAGTCTC 2023
QY 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTrp 668
Db 2024 ACGCGCTGACCAACACCTACAGATCGCCCTGGGTGGGTCA---GGCTCGAAGTGG 2080
QY 669 ProAlaValValValGlyThrIleGlyValThrGlyAla-----TyrArgSer 686
Db 2081 ---AACCTGTATGCTTGGCACC-----GGCCCGTCAAGGCGTCTCGCTCTACGCCAGT 2131
QY 687 AspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 706

Db 2132 GGAGACACGCGCGCTCTCTGGACGAGCATCTCCAGGGCTCCACGGGCTTCGGCTCCATCGAC 2191
QY 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGly 726
Db 2192 AGCACAAGTTCGCGGAGCGGACGCGGACGCGGCAAGTCTACGTGGGCAACACGCGC 2251
QY 727 ArgGlyIleValTyr-----GlyAspIleGlyAlaProSerGly 740
Db 2252 CGGGGCGTCTTTTACGCTCAGGGAACCGTCTCGCGCGGCGCACGCGCGG 2299

RESULT 7

US-10-156-761-2561
; Sequence 2561, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2561
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2217)

US-10-156-761-2561

Alignment Scores:
Pred. No.: 3,09e-125 Length: 2217
Score: 1405.00 Matches: 308
Percent Similarity: 54.57% Conservative: 104
Best Local Similarity: 40.79% Mismatches: 257
Query Match: 34.81% Indels: 86
DB: 15 Gaps: 21

US-09-917-376-3 (1-740) x US-10-156-761-2561 (1-2217)

QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
Db 124 TACCGTGGCGCAACGCCGCTCATCGGGGCGACCGGTCTGTCAACCGCGTCTTCCAC 183
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
Db 184 CCTCGGTACGCGGTCTCGCTACGCCGCGACCATCGCGCGCGCTACCGCTGGGAC 243
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
Db 244 GACCGCGCGCGCTGGACCGCGCTCATCGACCATCGGTGGGACGAGTGGAACTC 303
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
Db 304 CTGCGGCTGAGCGGATGCGTCTGACCCCGACCCCGACCGCGCGCTCTACCTCGCGGTG 363
QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
Db 364 GGACCTACCGCGAGTGTGGGCGGCAAC---GGCGCGGTCTCTCGCTCCGAGGACCGC 420
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
Db 421 GGCGCCACCTGGACCGCGACCGACCTGACCGTGAAGTCTCGCGCGCAACGAGGACGCGCGC 480

; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Alignment Scores: 1.22e-120 Length: 9025608
 Pred. No.: 1405.00 Matches: 308
 Score: 1405.00
 Percent Similarity: 54.57% Conservative: 104
 Best Local Similarity: 40.79% Mismatches: 257
 Query Match: 34.81% Indels: 86
 DB: 15 Gaps: 21

US-09-917-376-3 (1-740) x US-10-156-761-1 (1-9025608)

QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
 Db 3161882 TACCGTGGCGAAGCCGCTCATCGGGGACCGGCTTCGTACCGGCGTGTCTTCCAC 3161823
 QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
 Db 3161822 CCTCGGTACGGGCTTCGCTACGCCCGACCGACATCGGGCGGCTACCGCTGGAC 3161763
 QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
 Db 3161762 GACCGGGCGCGCTGGACCGCTCATCGACCATCGCTGGGACGACTGGAACCTC 3161703
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAlaAlaVal 85
 Db 3161702 CTGCGGCTGAGCGATGCGCTCGACCCCGACCGACCGCGGCTCTACCTCGCGCTG 3161643
 QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerAspGln 105
 Db 3161642 GGACCTACCGCCAGCTGGTGGCGGCAAC---GGCGGCTCTGCGCTCGGAGGACCGC 3161586
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLeuGlyGlyAsnMetProGlyArg 125
 Db 3161585 GCGCCACCTTGACCGCCGACCGACTGACCGTGAAGCTCGGCGGCAACGAGGACGCGCGC 3161526
 QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
 Db 3161525 GCGCGCGGTGAGCGACTCTCGTGGACCGCGCGACGACACCCCTCTGGTGGGCAACG 3161466
 QY 146 ProSerGlyIleGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
 Db 3161465 ---CGGACACGCGGCTGTCAAGTCGACGCGCGGCGGCGGCTTGGGAGCGCGGAC 3161409
 QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185
 Db 3161408 GCCTTCCG-----GGGAGGCGAAGTCC 3161385
 QY 186 AspIleGlnGlyValValTrpValAlaPheAspIleSerSerSerSerGlyGlnAla 205
 Db 3161384 TCGGGCAGGAGTCTGTTC-----CTCGTGGCGCC 3161352
 QY 206 SerIleThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220
 Db 3161351 GGGCGCACCTGTACCGCGCTGGGGTGCAGCGGACCGGACCTCGGGCACCGGCAACCTG 3161292
 QY 221 PheTrpSerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240
 Db 3161291 TACCGACGCGCCGAC---GGCAGACCTGGGGGCGCTCCCGCGGCGGCTCGGCGAC 3161235
 QY 241 -----IleProHisIleGlyValPheAspProValAsnHisValLeuTyrIleAla 257

Db 3161234 TCCGCCAAGTCCCGCTCGCGCGCGGTACGACACGACACCGCGCGAGCTGTACGTGACG 3161175
 QY 258 ThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpIlePheSerVal 277
 Db 3161174 TACGGCGACGACCCCGCGCGCGGTTCGACGCGAGCGTGCACAGCTGCGTACG 3161115
 QY 278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293
 Db 3161114 GCCACGGGACGTCGACCGAGGTACCCCGGTGAAGCGCGGCGGACGACGCGCGC 3161055
 QY 294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313
 Db 3161054 TCGGGCGACACCTTCGCTACGGCGGGTCCGCGCGGCGCGCGCGCGCGCGCGCTC 3160995
 QY 314 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 333
 Db 3160994 GTGCTCTCCACCAACACCGCTGGCGCGGACGCGGACGCGTCTTCGCTCCAGGACG 3160935
 QY 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353
 Db 3160934 GCGCGTACCTGGACGTCCTCAAGGAC-----GCCGCC 3160902
 QY 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373
 Db 3160901 GTGTTGACGCTGTCGAGACTCCCTTCTCGACTGGGCGGACGAC----- 3160857
 QY 374 ProSerProIleLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393
 Db 3160856 ---AAGCCGAAGTTCGCTGCTGGATCCAGGCGCTCGCGTCCGCTACGACTCCGAC 3160800
 QY 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413
 Db 3160799 CAGCTGCTGACGCGGACCGCGGACCTCTACGCHCCGCGACCTCAACGCGCTGG--- 3160743
 QY 414 SerGlyGlyGlnIleHisIleAlaProMetValIleGlyLeuGluThrAlaValAsn 433
 Db 3160742 -----GCACCGGATCCGCGGCTCGGAGGAGAGCGCGTGGC 3160704
 QY 434 AspLeuIleSerProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly 452
 Db 3160703 CAACTGATCTCGCCCGCGTCCGGGAGGACACCTGTATGACGCGACTCGGGGACATCG 3160644
 QY 453 GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal 471
 Db 3160643 GTGATGTACACGAGCGCTTCACGGCGTCTCCGTCGCGCGCATGGCGACGACCCCGTG 3160584
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491
 Db 3160583 TTCGGTGGCGAGCGGACTCGCGGCGCGCGCGCGCGCGCGGTATGTCTCGGACG 3160524
 QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
 Db 3160523 GGCTGGGGCGACACGCGCAACGCG-----GCGTACTCCGACGACGCGC 3160482
 QY 512 GlyIleAsnTrp-----PheGlnGlySerGluProGlyGlyValThrGlyGlyThr 529
 Db 3160481 GGGCGGACCTTGGCGCGCTTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3160422
 QY 530 ValAlaAlaSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545
 Db 3160421 ATGCCACCGAGTGCAGCG 3160365
 QY 546 ProGlyGlnProValValTyrAla-----ValGlyPheGlyAsnSerTrpAla 561
 Db 3160364 -----ACGACGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3160317
 QY 562 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProIleThr 581
 Db 3160316 GAGGTCTCTCTCTTCCGGAAGGGCGCGACACCGCTCGCGCGCGCGCGCGCGCGCGCG 3160257
 QY 582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThr 599

Db 3160256 TTCTACGCCTACGACTTCACAAATGGAACGCTATATACGCCAGCACTGACAGTGGCGCTTCG 3160197
 Qy 600 PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValAlaGlyValMetPheHisAla 619
 Db 3160196 TTACGGCCCTGCGGGCGGAGTGGCC---TCCGCGACACCGAGTTCAAGCTGGTTCGG 3160140
 Qy 620 ValProGlyIysGluGlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSer 638
 Db 3160139 GCGCGGGAGCAAGCGGCGACCTGTGGCTTCCGCGAAATGGAACGGGCTCTACCGGTCC 3160080
 Qy 639 ThrAsnGlyIysSerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658
 Db 3160079 ACCGACGGCGGGACACCTTCGCCAGGATGACAGCTGCTGGCGCTCTGACACCCCTCGC 3160020
 Qy 659 PheGlyIysSerAlaProGlySerSerTrpProAlaValPheValAlaGlyThrIleGly 678
 Db 3160019 TTCGCGCAAGGCGGCGGACGCGCGCGACCTACCGCGGATCTACAGGTGGCTCGACGGAG 3159960
 Qy 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp 698
 Db 3159959 ACCATCACCGCGCTACCGCTCCGACGACGCGCGCAGACATGGTCCGGATCAACGAC 3159900
 Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718
 Db 3159899 GACGCCACCAAGTGGGGTGGATCGCGGAGCGCGTGGTGAACCCGCGATCCACCGC 3159840
 Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733
 Db 3159839 CGGGTCTACCTCGCCACCAACGCGCGCGCATCCAGTACGGGGAG 3159795

RESULT 9

US-09-927-827-21
 ; Sequence 21, Application US/09927827
 ; Publication No. US20030036176A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bower, Stanley G.
 ; APPLICANT: Ramseier, Thomas M.
 ; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
 ; FILE REFERENCE: 38-10(15824)B
 ; CURRENT APPLICATION NUMBER: US/09/927,827
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: US 60/279,493
 ; PRIOR FILING DATE: 2001-03-28
 ; NUMBER OF SEQ ID NOS: 69
 ; SEQ ID NO 21
 ; LENGTH: 3668
 ; TYPE: DNA
 ; ORGANISM: Xanthomonas campestris
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1001)..(2668)
 US-09-927-827-21

Alignment Scores:
 Pred. No.: 2,866-99 Length: 3668
 Score: 1139.00 Matches: 267
 Percent Similarity: 50.85% Conservative: 120
 Best Local Similarity: 35.09% Mismatches: 295
 Query Match: 28.22% Indels: 79
 DB: 10 Gaps: 21

US-09-917-376-3 (1-740) x US-09-927-827-21 (1-3668)

Qy 1 AlaThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20
 Db 1217 GCCAGTCCGGCCCTACAGTGGCGCAGTGTGCGCATTTGGCGGTGGCGCTTTGTCAAC 1276
 Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40
 Db 1277 GGTGTGCTGTTTCATCCCGCCCGAAGCGTGGTCTGGGCTATGCGCGCACCGATGTGGTGGC 1336
 Qy 41 MetTyrArgTTrpAspAlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrpValGlyTrp 60

Db 1337 CGGTACCCCTGGAGTGCAGCGCGCAGCAGTGGACGCGCCTGACCCAGCTGTTGGGGCGCT 1396
 Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
 Db 1397 GACGACTGGAACCTGATGGCATCGACGATTCGCGTGCAGCCCGCCGATGCCGATGCG 1456
 Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
 Db 1457 CTGTATCTGGCGCGCGCACCTATATGCATGAA--CGCGCGGCGCACTCCGCGAGTGTG 1513
 Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheIysLeuGly 120
 Db 1514 CGCTGTTCAACCGCGCGCGCATGTCGACGTTCGAGCGTGGCGACTGCGCTTTAAGCTGGGTGT 1573
 Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
 Db 1574 AACAGCTGGGCGCGCCCAATGGCAGCGGCTGGCGGTGGACCCGACGATGGCCGCGTG 1633
 Qy 141 LeuTyrPheGlyAlaProSerGlyIysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
 Db 1634 CTGCTGCTGGGCTCGCGCGATGCC--GGCCTGTGGCGTAGCGACGATCGCGCGCGCAC 1690
 Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
 Db 1691 TGGGCGAAGGTGGCGCTGTTCCGAGCGCGCGCTGGCGCGTCCACCGCGCGCAATCAT 1750
 Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200
 Db 1751 GTTGGCGCGGACGACGGGTG--GGATCGCTTTGTCTGTTCGACGCGACCATGGC 1807
 Qy 201 SerLeuGlnAlaSerIysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
 Db 1808 RACAAACGGCTCGCCCAACCGCGCATCTACGTGGCGTGTCCACCGACAGACCGCTG 1867
 Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240
 Db 1868 TATGTGTCCGAAGATCGCGCGCGAGTTGGGACCGCTGGCGGCAACCGCGCGCTG 1927
 Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
 Db 1928 CGCCCGAGCCACATGGCGCGCGCAGCGATGGGCAC--TGGTATCTGAGCTATGGCGAC 1984
 Qy 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280
 Db 1985 CAGCCCGCGCGACCTGATGCGCGGGAGCCTTGTGGAAATTCACCGCGCGCACAGGG 2044
 Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
 Db 2045 CGCTGGCGTGAATCAGCCCGATTCGCGAG--CCAGCCAGTGGCGCATGGATTCGGGTGG 2101
 Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
 Db 2102 GGTGGGTGGCGGTGGATCCGCAACATCCGAGGTGCTGTGGCCAGCACTTCCGCGCT 2161
 Qy 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340
 Db 2162 CGCACCGCGCGCAGCGCTGTATCGACGCTGATGGTGGCAAGCACTGGACCGCGTGG 2221
 Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
 Db 2222 -----TTGGCGGATGCGGTGTTCGATCAGACGCGCGCG 2254
 Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrp 380
 Db 2255 CGGTG--ACCGCATATGCCACGCGCGAC-----TGG 2284
 Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
 Db 2285 ATGGGG--CGCTGGCGATCGATCCGTTCCAGCGCAACCATCGCTGTTCGTGACCGCG 2341
 Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
 Db 2342 TACGGCATCTGGGCTCGCGCAATCTG-----CAGGATTTCCGC 2380


```
QY 421 AlaPro-----MetValLysGlyLeuGlucluThrAlaVal 432
DB 2381 GCACCGCAGCGCCGCTCGAGTGGTTCAGAGACCGTGGGTGGAAGAACCGTGGCG 2440
QY 433 AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 452
DB 2441 CTGACCTGCTCAGCCGATGCTGGCGGCATCTGCTCAGCGCGCTCGGCATATCGAC 2500
QY 453 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 472
DB 2501 GGCTCCCGCATGACGACTGACCGCGTG---CAGTTGCAGTACGCGCGCCCGCGGTG 2557
QY 473 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 492
DB 2558 ACCAATGGCAAGCATGATCCCGCCGACGCGCGCGCGCGCGCGCGCGCGCGT 2617
QY 493 SerPhe-AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyG1 512
DB 2618 ACCGTGGCGCGCGCGCGCAACAGAAATCCGCGCGC---TGTATTACGCGTATGGCGG 2674
QY 512 yLysenTrpPheGln---GlySerGluProGlyValThrThrGlyGlyThrValAl 531
DB 2675 CAAGCAATGACTGCTGCTGGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGTTC 2734
QY 531 aLaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVa 551
DB 2735 CATTTGTGCGATGCGCGCGCAGTGTGTGGCACCGCGCGCGCGCGCGCAAT-----TG 2788
QY 551 lYrAlaValGlyPheGlyAsnSerTrpAlaLaSerGlnGlyValProAlaAsnAlaG1 571
DB 2789 GCGCAGTCCGACTTCCGCGCGCGAGTGGCAGGTGTGAACGCCCTGCGCAACCCCGCGT 2848
QY 571 nileAsrSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyTh 589
DB 2849 GGTGTGACCGATCGGTGTGACGCGCGCGCTGTTATGGTGGTGTGCGCGCGCGCGCA 2908
QY 589 rPheTyrArGserThrAspGlyValThrPheGlnProValAlaAlaGlyLeu----- 607
DB 2909 GCTGTACGAGACACCGATGCGCGCGCAGTTTCCGT-----GCGACCGGTGTACAGT 2962
QY 608 -----ProSerSerGlyAlaVa 613
DB 2963 GGGCAGCCCGCGCGCGATGAGCGCACCCGCGCGCAGTGCCTCGGACCCCGTGGCGCGC 3022
QY 613 lGlyValMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaLaSerSe 633
DB 3023 AGCGGTGGTGTATCTGCGCGACGCCCGCGCAG-----ProSerSerGlyAlaVa 613
QY 633 rGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSe 653
DB 3054 -GGTGTGATGCGGTGCAGGACGCT-----GCGTGCAGGTGCTCTCACACCGCGACGA 3106
QY 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVa 673
DB 3107 AGCACGCTGCTGGCATCGCAAGCGCTGCGCGCGCGTGGCGCGCGTGGCGCGGTATCT 3166
QY 673 lValGlyThrIleGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTr 693
DB 3167 GGGCGGCGCGGTGACGGCGGTGATGGCGTGTTCCTGCTCGACGATGGCGCGCGCGATG 3226
QY 693 pValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAs 713
DB 3227 GCAGCGCATCAGCATGACGCGCACCGCTTCGGCGC---CCGTACAGCGTGAACCGTGA 3283
QY 713 pHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAs 733
DB 3284 TCCGCGCATTCGCGCGGTGTGTACTTGCACCGCGCGCGCGCGCGCATTTTCTACGCGCA 3343
QY 733 p 733
DB 3344 T 3344
```

RESULT 10

```
US-10-395-241-13
; Sequence 13, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasuishi
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2367
; TYPE: DNA
; ORGANISM: Geotrichum sp. M128
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (1)..(2367)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2367)
; US-10-395-241-13
```

Alignment Scores:

Pred. No.:	2,18e-97	Length:	2367
Score:	1117.00	Matches:	281
Percent Similarity:	50.37%	Conservative:	124
Best Local Similarity:	34.95%	Mismatches:	279
Query Match:	27.68%	Indels:	120
DB:	18	Gaps:	31

US-09-917-376-3 (1-740) x US-10-395-241-13 (1-2367)

```
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
DB 10 TAGCAGTTCAAGAATGTCGCGATCGCGCGCGGGGTACATTACCGGGATTGTCGCGCAC 69
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
DB 70 CCAAGACCAAGACCTGCTGTACGCGCGCGCGACATTCGCGCGCGGTACCGTGGGAC 129
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
DB 130 GCAGGCACGTCCAAGTGGATCCGCTCAACGACTTTATCGAGGCGCGACATGAACATT 189
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
DB 190 ATGGGCACCGAGTGCATCGCGCTCGACCCCAACACCCGACAGCTGTACCTCGCGCAG 249
QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104
DB 250 GGGCGTATGTCGGCGACGAGTGG-----CGCGGTTCTATGTCTCGGAAC 297
QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124
DB 298 CGCGGCGAGTCTGTTTACATCTACGAGTCCCGTTCCTCGATGGCGCGCACATGGGA 357
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144
DB 358 CGCAACAAATGGCGAGCGCTCGCTGTCAACCCGTTCACTCAACGAGGTCTGTATGGT 417
QY 145 AlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
DB 418 ACOCGTACA---GAGGGTATCTGGAGAGTTCGAGACCGCGCGCAAGATCTGGACAAACGTC 474
QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184
DB 475 ACGTCCATCCGAGACGGTTC-----ACCAACGGTATCGGATACACG 516
```



```
; LENGTH: 2646
; TYPE: DNA
; ORGANISM: Geotrichum sp. M128
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)..(2558)
US-10-395-241-11

Alignment Scores:
Pred. No.: 2,516-97 Length: 2646
Score: 1117.00 Matches: 281
Percent Similarity: 50.37% Conservative: 124
Best Local Similarity: 34.95% Mismatches: 279
Query Match: 27.68% Indels: 120
DB: 18 Gaps: 31

US-09-917-376-3 (1-740) x US-10-395-241-11 (1-2646)
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyLeValPheAsn 25
DB 198 TAGGAGTTCAAGAAATGTCGATCGCGCGCGGGGTACATTACCGGGATGTCGGCAC 257
QY 26 GluGlyAlaProGlyLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45
DB 258 CCAGAACCAAGACCTGCTGACGCGCGCACGGACATTGGCGCGGTACCGTGGGAC 317
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
DB 318 GCAGGCACGCTCCAGTGGATCCGCTCAACGACTTTATCGAGCGCGGACATGAACATT 377
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
DB 378 ATGGGCACCGAGTCGATCGCTGGACCCCAACACCCGACAGCGGTGATCTCGGCGAG 437
QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104
DB 438 GGGCGGTATGTCGGCGACAGTGG-----GCGCGGTTCATGTGTCGAGAC 485
QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGlyAsnMetProGly 124
DB 486 CGCGCGCAGTCGTTTACATCTACGAGTCGCGGTTCGATGGCGCCCAACGACATGGGA 545
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144
DB 546 CGCAACAATGGCGAGCGCTCGTGTCAACCCGTTCAACTCGAACCGAGGTCTGGATGGGT 605
QY 145 AlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
DB 606 ACGCGTACAA---GAGGGTATCTGGAAGAGTTCGGACCGCGGCCAGACCTTGGACAAACGTC 662
QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184
DB 663 ACCTCCATCCGAGCCCGTTC-----ACCAACGGTATCGGATACACG 704
QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204
DB 705 TCG-----GTCAUTTTTCGACCCC-----GAA 725
QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224
DB 726 CGTAATGGCACCATCTACGCGAGCGGACTGCGCCCGAGGGC---ATGTACGTACGCGAC 782
QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle-----241
DB 783 GACGGCGGTCTCTGTTGGGAGCCAGTGGCGGGCCAGCGCTCCAGTGGCTCAACAGGACC 842
QY 242 -----ProHisLysGlyVal 246
DB 843 ACGGGCGGTTCGCGACAAAGACCGCGTCGATCGCGCGCGACGCCCATGAAGTCGT 902
QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266
DB 903 CTCACCCCC-----AACTTCCTCTACGTGACTTACCGCGACTACCCCTGGTCCATGGGGC 956
```

```
QY 267 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle--- 285
DB 957 GTCACGTCGCGGAAAGTCTGGCGCAGAACCGCACCTCGGCGCCTGGGACGACATTACT 1016
QY 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
DB 1017 CCCCCGCTCGGCAACTCGTCGCTGCCCGTCAACACACGACAGCTTCCCTCGCGGCGGA 1076
QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
DB 1077 TTTTGGGCTCTCAGCGTCAGCGCACCAACCCCAACCGTCTCTGCTCATCACC---CTC 1133
QY 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336
DB 1134 GACCGGACCCCGGACCCCGCTCGACAGCATCTACTCTCAACCGATCGCGGCGGAC 1193
QY 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347
DB 1194 TGGAGGAGCTCACCCAGCTCTCGTCCCGTCCAACTCGAAGGTAACCTGGGCGCACCCG 1253
QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 367
DB 1254 ACTAAGCGCGCGGTAC---AAGGACGCGACCGCTGTTCCGTGGCTCGACTTCAACAAC 1310
QY 368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379
DB 1311 GGTCCCACTGGGGGGATACGGTGGCGCGCACGGTACGCCCGCGCTCACCAAGTTTGGC 1370
QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399
DB 1371 TGGTGGATGAGCGTGTGTTATCGATCCGTTCAACCCCGACGACCTGTATGTACGGCAG 1430
QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 415
DB 1431 GGGCGCACCATCTGGCGCGACCGACACGCTCTCCGCTGTCGAGAGGACTGG----- 1481
QY 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 431
DB 1482 -----CGCGCGAGCTGCTACCTCCAGTCCAGTACCGTATCGAGGAGATGGC 1526
QY 432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451
DB 1527 ATCTCTGCTCGCTCGCTCGCCAAAGAGCGCGCGGCTCTCTGTCGGGCATCGGTGACATT 1586
QY 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471
DB 1587 AGCGGCATGAAGCACGACGACCTCACCAAG---CCCCAGAAAGATGTTTGGTGGCCCCCAG 1643
QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491
DB 1644 TTTCTCAACCTCGACAGCATCGACGCTCGCGGCAACTTCCCAACGTTGCTCGCGGCC 1703
QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511
DB 1704 GGATCTCGGACACAGTACGACGCGCGCGCGCGGTGGTGGTACGCGACTGACGCG 1763
QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThr----- 526
DB 1764 GGAGACGCTGGACCATCTTCCCTACCTGCTCTCTGCGCATGAACGCGGACCTACCCAG 1823
QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 544
DB 1824 GGCAGCACGATTCAGTTCGACGCGGCGGCGGACGATCGTGTGGTGCACCAAGCTTGAC 1883
QY 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564
DB 1884 GAGCAGGCTCGGGACCGTGTACTCTCCACGACTATGGCAAGAGCTGTCT------ 1934
QY 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578
DB 1935 ---GTTCCCGCTGGCGACCTGAAGGCCAGACTGCCCAATGTGCTCTCGGACAAAGGTCCAG 1991
```


Qy 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 598
 Db 1992 GATGGACAGTTCTACGCTACCGATGGCGCAAGTTCTTCGTCTCGACCGCGGGAAG 2051
 Qy 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618
 Db 2052 TCGTATGCGCGCAAGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 2105
 Qy 619 AlaValProGlyLysGlyGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 637
 Db 2106 GTGAACCCCTGGGTGGCGCGGAGCTCTGGGTGCTGTTCGGAGGGCGGTCTCTTCAC 2165
 Qy 638 SerThrAsnGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656
 Db 2166 TCGACCGACTTGGCGCTGCTTACGAGGCTAGTACCGCAACGCGACCTCGTGAGC 2225
 Qy 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672
 Db 2226 GTCGGCGCGCCCAAGTCCAGTCGCGCAAGAGGCTAGCGCGCCCTCCGCGGTCTTC 2285
 Qy 673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 690
 Db 2286 ATCTGGGCGCACCGACAGCTGGAGCGCATCGGCTGTACCGCTCGACGACACCGC 2345
 Qy 691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710
 Db 2346 AGCACCTGGACGCGCTCAATGACGAGGACACACTACTCGGGC---CCACCATGATC 2402
 Qy 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyIleVal 730
 Db 2403 GAGGCGCGACCCCAAGTCTACGGCGCGGTGTATCTAGGACGACGCGCGGTATCTGTG 2462
 Qy 731 TyrGlyAspIle 734
 Db 2463 TACGCGGACCTT 2474

RESULT 12
 US-10-395-241-17
 ; Sequence 17, Application US/10395241
 ; Publication No. US20040038367A1
 ; GENERAL INFORMATION:
 ; APPLICANT: YAOI, Katsuro
 ; APPLICANT: MITSUTSUSHI, Yasushi
 ; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
 ; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
 ; FILE REFERENCE: Q73756
 ; CURRENT APPLICATION NUMBER: US/10/395,241
 ; PRIOR FILING DATE: 2003-03-25
 ; PRIOR APPLICATION NUMBER: JP 2002-83433
 ; PRIOR FILING DATE: 2002-03-25
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 17
 ; LENGTH: 2481
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Xyloglycan Oligosaccharide-Degrading Enzyme with Histidine Tag
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2481)
 US-10-395-241-17

Alignment Scores:
 Pred. No.: 4,51e-97 Length: 2481
 Score: 1114.00 Matches: 281
 Percent Similarity: 50.25% Conservative: 123
 Best Local Similarity: 34.95% Mismatches: 280
 Query Match: 27.60% Indels: 120
 DB: 18 Gaps: 31

US-09-917-376-3 (1-740) x US-10-395-241-17 (1-2481)

Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25
 Db 13 TACGAGTTTCAAGAATGTCGATCGCGCGCGGGTACATATACCGGGATTCGCGCAC 72
 Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45
 Db 73 CCAAAGACCAAGGACCTGCTGTACGCGCGCACGACATTGGCGCGCGCTACCGCTGGGAC 132
 Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65
 Db 133 GCAGCGACGTCCAAGTGGATCCCGCTCAACGACTTATATCGAGCGCGCACATGACAT 192
 Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaVal 85
 Db 193 ATGGGACCGAGTTCGATCGCGTGGACCCCAACACCCCGACAGGCTGTACCTCGCGCAG 252
 Qy 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerAsp 104
 Db 253 GGGCGCTATGTCGCGCAGCAGTGG-----GCGGCGTTCATATGTGTCGGAAGAC 300
 Qy 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124
 Db 301 CGCGCGCAGTCGTTTCAATCTACGAGTCGCGCTCCGATGGCGGCCAACACATGGGA 360
 Qy 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144
 Db 361 CGCAACAATGGCGAGCGCTCGCTGTCAACCGCTTCAACTCGAACGAGGCTCTGGATGGT 420
 Qy 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164
 Db 421 ACGGGTACAC---GAGGGTATCTGGAGAGTTCGGACCGCGCCCAAGCCTGGCAACACGTC 477
 Qy 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184
 Db 478 ACCTCATCCCGAGCGCTTC-----ACCAACGATATCGGATACACG 519
 Qy 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204
 Db 520 TCG-----GTCAATTTTCGACCCC-----GAA 540
 Qy 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224
 Db 541 CGTAATGACCATCTACGCGAGCGCGACTGCCCGCGAGGGC---ATGTACGCTCAGCAC 597
 Qy 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241
 Db 598 GACGCGGTGTCTCGTGGAGCCAGTGGCGGCGCGCTCCAGCTGGCTCAACAGGACC 657
 Qy 242 -----ProHisLysGlyVal 246
 Db 658 ACGGCGCGTTCCTCCGACAGAAGCCGCTCGATCGCGCGCGCCATGAAAGTCGCT 717
 Qy 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266
 Db 718 CTCACCCCC---AACTTCCTTACGTGACTTACGCGCGACTCCCTGCTCATCGGGC 771
 Qy 267 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle--- 285
 Db 772 GTCACGTTCCGCAAGTTCGCGCGCAGAACCGCACTCGGGCGCTCGGGACGACATTA 831
 Qy 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
 Db 832 CCCCGCTCGGCAACTCGTCGCTCCCGGTACACCAACACGACGCTTCCTCGGGCGGA 891
 Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
 Db 892 TTTTTCGCTCTCAGCGTCGACGCGACCAACCCCAACCGCTCTCGTCTCATCACC---CTC 948
 Qy 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyAlaThr 336
 Db 949 GACCGCGACCGCGGACCGCGCTCGACAGCATCTACCTCAACCGATCGCGCGGACC 1008
 Qy 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347

Db 542 CGGCTGCTGAGTACCTGGCGCAGTGCCTACCCCTAACGACTATTATACGCCCGCAGCGTGTTC 601
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85
Db 602 TGGGGCATCGACGCTGTTGGCTTTGATCCGCGAGCAGATCAAAAGGTGTATGCCGCGAGTG 661
Qy 85 ----- 85
Db 662 ACCCGTAGCTGCAGCAACCGGAACCTAGGCGTCTCTGCTAGTGTTCACATACGCCGTAC 721
Qy 86 GlyMetTyrThrAsnSerTrpAsp----- 93
Db 722 GGCATGTATACGAACAGCTGGTC-TGTGATGTCTCTCAGATCTAGACCTATGATTTGGACG 780
Qy 93 ----- 93
Db 781 GCCGTACATATGCTTGTTCGACAGACACTACAGAGAGTCTAGATCTGGATACTAACCTGC 840
Qy 94 -----ProAsnAspGlyAlaIleLeuArgSerSerAspGln 105
Db 841 CTGACATGGCCATATATAGGGATCCGAGTAATGAGCCATCATTCGCTCGTCAGACCGC 900
Qy 105 ----- 105
Db 901 GACTGTAAACGGTATATATCCTAGGCTCATTAACCTCGTAGTAAGCGAGCAGTCTGGCG 960
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125
Db 961 GCGCAACGTTGCTCTTCAACCACTTGCCCTTCAAAAGTCGGGGGTAAACATGCCAGGACGC 1020
Qy 125 ----- 125
Db 1021 CCGCGTTGCACAGGAAGTGTGAACGGGAAGTTTCAGCCCCATTCATGCTACGGTCTCGCG 1080
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145
Db 1081 GGAGCCGGAGAGCGTCTGGCTGTGCATCCGCCAACTCCACATCACTACTTCTTGGTGTCT 1140
Qy 146 Pro----- 146
Db 1141 CCTCGGCTCTCGCAGACCGCAGCTAGGCGGTTGAGTTGTAGTAGTGAACACCGA 1200
Qy 147 ---SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet--- 164
Db 1201 CGCTCAGGAACCGCCTCTGGAAGTCTACGACGGCGCGTGAACCTTTTCCAAGTCTCG 1260
Qy 165 -----Thr 165
Db 1261 CGGAGTCTTTGCGGAGACCTTCAGATGCTGCGCGCGCACTGGAAAGGTTCCAGAGC 1320
Qy 166 AsnProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln--- 184
Db 1321 TCGTTACGGCAACTGGGAGCTACATCCAGACCGCGAGTGATTCACACGGCTACAACAGC 1380
Qy 185 -----Ser 185
Db 1381 AGCAAGTGCCTTGACCCCTGCATGTAGGGTCTGGGCTCACTAAGGTTGCCGATGTTGTCG 1440
Qy 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205
Db 1441 GACAAGCAAGCACTCATGTGGTTACGTTTCGACTCAACACAGCAGCACCGCGGGGAGCC 1500
Qy 205 ----- 205
Db 1501 CTGTTGCTTCCTGAGTACACCAATGCAAGCTGAGTTGGTGGTCTCGTGGCCCCCTCGG 1560
Qy 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp----- 222
Db 1561 ACGTCTCGTATCTTGTGGCAGCGCTGATAACATCACTACCTGCTTCAGTCTATGTGAGCAGC 1620
Qy 223 -----SerArg 224
|||

Db 1621 TGCAGAGCATAGAAACAACACCGTGGCCGACTATTGTAGTGACGAAGTCAGATACACTCGTGC 1680
Qy 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHis--- 243
Db 1681 AATGCCGCTCCAGCTGGAGTGTCTGATCCGGGCGACCGAGGAATACTTTCTCTACAAG 1740
Qy 244 -----LysGlyValPhe 247
Db 1741 TTACGGCCGAGGTGCACCTCACGACATGSCCCCGTCGGTCCCTTTATGAAGAGAGTGTTC 1800
Qy 248 -----AspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyPro 264
Db 1801 GCGAAACTGCAAGCCAGCAGAGAAGCCTTGTATCTGACCTATTCCGATGGCAGGCGCG 1860
Qy 264 ----- 264
Db 1861 CGTTTGAAGTGGTGTCTCTTCGGAACATAGACTGGATGAAGGTACGGTGTCCCGGC 1920
Qy 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284
Db 1921 TATGATGGCACACACTTGGCTCAGTGTGGAGTACGACATTCAGGGGGAACCTTGGAAAGAC 1980
Qy 285 IleSerProValProSerThrAspThr----- 293
Db 1981 ATACTACCGTGTGAACCGAGTCACACCTCCATCTGTAACTCCCTTTGAACCTTTCTG 2040
Qy 294 -----AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp--- 306
Db 2041 ATCACCCTGCTCTCGGATCAGATCTATCTTTGGCTTTGGCGGCTTGGCTTCGATTTG 2100
Qy 307 -----Arg 307
Db 2101 TAGTGGGACAGAGACCTAGTCTAGATATGAACCCGGAACCCGCGAGCTAAAC 2160
Qy 308 GlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr----- 325
Db 2161 CAAAGCCAGGAACCCCTGTTGCTTCTTTGAACTCTTGGTGGCAGATGCTCAGCTG 2220
Qy 326 -----IleIle 327
Db 2221 GTTTTCGGTCTTGGAAACAACAAGAAACTGTGAGAACCAACCCGCTCTACGAGTCGAC 2280
Qy 328 PheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAspThrTrpSerTrpPro 347
Db 2281 TTTCCGTGACCGACTCTGGGACACATGAGCCCGCATCTGGCGTGGGCGACTATCCG 2340
Qy 348 AsnArgSer---LeuArgTyrValLeuAspIleSerAlaGluPro----- 361
Db 2341 AAAGCCAGCTGGCTGAGACC-CTGTTGTACCTCGGCTAGACCCGCAACCCGCTCGATAGG 2399
Qy 362 -----Trp 362
Db 2400 CACTGAGACCTATTACTACAGCATCTCAGTGAGTCACTTTTACGATCCGATCGAATGG 2459
Qy 363 LeuThrPhe-----GlyValGlnProAsn----- 370
Db 2460 CTGACTCTGGATAATGATGTCTGAGTCACTCAGTGAGATTTGTAGGTACGCTTACC 2519
Qy 371 -----ProProValPro----- 374
Db 2520 GACTGACTCTGCTTCAAGACTCCCAAGCACCCTGGATCAAGAAACAACATTTATCGATGTG 2579
Qy 374 ----- 374
Db 2580 ATGACTGAGACGAAGTTCTGAGGTTTCTGTCGACCTAGTTCCTTGTGTAATAGCTACAC 2639
Qy 375 -----SerPro-----LysLeuGlyTrpMetAspGluAlaMet 385
Db 2640 TCGACGAGTCAACCGTCCGATGGTCTCATCAAGCGCTCTGGCTGGATGATTGAGTCTCTC 2699
Qy 386 Ala----- 386
Db 2700 GGCTCGCTCAGTGGCAGGCTACCAGAGTAGTTTCGGGAGCCGACCTACTAATCTCAGAGAG 2759

QY 387 ---IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405
 Db 2760 CAGATTGACCCAAACCGACAGCAACACCTGGCTCTACGGCACCGGAATGACAATCTTTGGC 2819
 QY 406 Thr----- 406
 Db 2820 GTCTAACTGGGTGGCTGCTGTTGGTACCGAGATGCGGTGGCTTACTGTTAGAAACG 2879
 QY 407 ---AsnAspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLys 425
 Db 2880 CGCCACGATCTCACCAACTGGGACACCGCGCACCAATGTGTCAATCCAATCACTGGCGAGAC 2939
 QY 426 Gly----- 426
 Db 2940 GCGGTGCTAGAGTGGTTGACCTGTGCGCGGTGTACACAGTTAGTTAGTACGCTCTG 2999
 QY 427 ---LeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIle 445
 Db 3000 CGATCGAGGAATTCCTCGTCCAGACCTGGCTCTGCACCCCGGCGGAAGCGAGTATTG 3059
 QY 446 Ser----- 446
 Db 3060 GCGTAGCTCTTAAGAGGCAGGTCTCTGGACCGGAGAGTGGCGCGCTTCGCTCGATAAC 3119
 QY 447 ---AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal----- 462
 Db 3120 CCGCAGTCCGAGACGACAAACGGCTTCACCTTTGGCCAGCAGAAACGACCTCGGACATCG 3179
 QY 462 ----- 462
 Db 3180 CGGCGTACGCTCTGCTGTTGCCGAAGTGGAAACGGTCTCTTGTGAGCGCTCTAGC 3239
 QY 463 ProSerThrIlePheThrSerProValPheThrGlyThrSerValAspTyrAla--- 481
 Db 3240 GCGCAGCGGTCTGGGCAACGCCACATGGGCCACCTCGACGCGTCTGACTACGCGCG 3299
 QY 482 -----Glu 482
 Db 3300 AGCGTCTGCAGACCGCTTGGGTGTACCCGGTGGAGTGTCTCGCAGCTGATGCGGCC 3359
 QY 483 LeuAsnProSerIleIleValArgAlaGlySer----- 493
 Db 3360 TACTCGTCAAGAGCGTCTCGCGTCCGCAACACCGCGGACGCAACAGGTGGCCATC 3419
 QY 494 ---PheAspProSerSerGlnProAsnAspArgHisValAlaPhe----- 507
 Db 3420 TTGAGCCAGTTCTCGCAGCAGGGCGAGCCGTTGTGGCGCGCTGTGTCCACCGGTAG 3479
 QY 508 SerThrAspGlyLysAsnTrpPheGlnGlySerGluProGlyValThrThrGly 527
 Db 3480 ACGTCCGACGGCGCGCAGCTGGACATCGACTACGCGCGCCGACACGTCATGAACGGC 3539
 QY 528 Gly----- 528
 Db 3540 GCGAGGCTGCGCGCGCTGCACCTCGTAGCTGATGCGCGCGCTGTGCAAGTACTTGC 3599
 QY 529 ---ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGly 547
 Db 3600 CGCAGGTGGCTTATTCGGCCGACCGCGACATGATCTCTGTGTGACCGCTCGTCCGGC 3659
 QY 548 -----GlnPro 549
 Db 3660 GCGTGCCACCGGATAGCGGCTGCGCTGTGTAGGAGACCACTGTGGCGGAGCAGCGC 3719
 QY 550 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsn 569
 Db 3720 CTGACGGCTCG---CAGTTCCAGGCGAGTTTGGCTCTCGAGCTTCGCGCGCGCGC 3776
 QY 570 AlaGln----- 571
 Db 3777 GCCGAGTCCGAGCGCTCAAGGTCCCGTGGAAACGGAGGACGAGCTCGGACGGCGCGCC 3836

QY 572 ---IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThr 589
 Db 3837 CGGCTCATCGCTCGCAAGAAGACCAACAGCGTCTTCTACCGCGCTCCGATCGAC 3896
 QY 590 Phe----- 590
 Db 3897 TTTTAGTAGCGAGCCTGTTCTTCTGTTGTCCGAGAAGATGCGCGAGCCTAGCTGG 3956
 QY 591 ---TyrArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuProSer 609
 Db 3957 AAAACGTCAGCAAGACACCGGACAGCTTC-----ACGCGCGG---CCCAAG 4004
 QY 610 SerGlyAlaValGlyValMet----- 616
 Db 4005 CTGGGACGCGCAGGATGCGTCTGTCGCGTGTGCGAAGTGTGCGCGCGCGGCTTC 4064
 QY 617 -----PheHisAlaValProGlyLysGluGlyAspLeu 627
 Db 4065 GACCCGTCGCTCCCTCGATCCGGGATATCGTCTCACCCGACACCGCGGCGACGTTG 4124
 QY 628 TrpLeuAlaAlaSerSer----- 633
 Db 4125 TATGCTCGACCGACCGGTAGGCCCTATAGCGAGTGGGTGTGCGCGCGCGCTGCAAC 4184
 QY 634 -----GlyLeuTyrHisSerThrAsnGlyGlySerSer----- 644
 Db 4185 ATACAGAGTGGTGTCTCGCATATTCGCTCCACAGACTCGGCGACACCTTTGGCCAA 4244
 QY 644 ----- 644
 Db 4245 GTCTCCACCGCCTGAAGCGGTATAGGCGGAGGTGTCTGAGCCGCTCTGGAACCGGTT 4304
 QY 645 -----TrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662
 Db 4305 CAGAGTGGCGGACTCCAACTACAGATCGCCCTGGTGGGTGGCTCAGGCTCGAAC 4364
 QY 663 -----AlaProGlySerSerTyr 668
 Db 4365 TGGAACTGTATCGCTGGTGTGGTGTAGGGGACCCACACCGAGTCCGAGCTTG 4424
 QY 669 ProAlaValPheValValGlyThrIleGlyValThrGlyAla-----TyrArgSer 686
 Db 4425 ACCTTGCATACCGATCGGCACC-----GGCCCGTCAGGGGTCTCGCTCTACGCCAGT 4478
 QY 687 AspAspCysGlyThrThrTrpVal----- 694
 Db 4479 GGAGACAGCGCGCTCTGGAAGCGGTGGCGCGGAGTCCCGGCGGAGATCGGTCA 4538
 QY 695 -----LeuIleAsnAspAspGlnHisGlnTyrGly----- 704
 Db 4539 CCTCTGTGCGCGGAGGACCTCGACATCCAGGCTCCAGGCTTCGGCTCCATCGAC 4598
 QY 704 ----- 704
 Db 4599 AGCACCAAGTCCCGGCGCGCTGTAGTCTCCGAGGGTCCCGAAGCGGAGGTAGCTG 4658
 QY 705 AsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThr 724
 Db 4659 TCGTGGTTCAGCGCGCGCTCGCGCAGCACCGCGCGG-----CAAGTCTACGTGGGACC 4712
 QY 725 AsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738
 Db 4713 AACGGCGGGGCGTCTTTTACGTCTAGGCGTCTAGGCGTGTGCGGCGGCC 4754

RESULT 14
 US-10-653-047-7511
 ; Sequence 7511, Application US/10653047
 ; Publication No. US2004029367A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Randy M. Berka
 ; APPLICANT: Michael W. Rey
 ; APPLICANT: Jeffrey R. Shuster
 ; APPLICANT: Sakari Kauppinen

APPLICANT: Ib Groth Clausen
APPLICANT: Peter Bjarke Olsen
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
TITLE OF INVENTION: Expression
FILE REFERENCE: 5849.200-US
CURRENT APPLICATION NUMBER: US/10/653,047
CURRENT FILING DATE: 2003-08-29
PRIOR APPLICATION NUMBER: US/09/533,559
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/273,623
PRIOR FILING DATE: 1999-03-22
NUMBER OF SEQ ID NOS: 7860
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7511
LENGTH: 1103
TYPE: DNA
ORGANISM: Tricoderma reesei
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(1103)
OTHER INFORMATION: n = A,T,C or G
US-10-653-047-7511

Alignment Scores:
Pred. No.: 7.64e-28 Length: 1103
Score: 392.50 Matches: 114
Percent Similarity: 45.05% Conservatives: 36
Best Local Similarity: 34.23% Mismatches: 125
Query Match: 9.72% Indels: 58
DB: 20 Gaps: 10

US-09-917-376-3 (1-740) x US-10-653-047-7511 (1-1103)

Qy 302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 321
Db 3 GGCCTGCGCTCGATTGCAAGACCGACCTGTTGTTGTTCTTGAACCTCTGG 62
Qy 322 TrpProAspThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrp 341
Db 63 TGGCCAGATGCTAGCTGTTTCGTCGACCGACTCTGGGCAACATCGAGCCGCTGG 122
Qy 342 AspTrpThrSerTrpProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 361
Db 123 GCGTGGGAGCTATCCAGCTGAGACCTATTACTACAGCATCTCACTCCCAAGCACCG 182
Qy 362 TrpLeuThrPheGly---ValGlnProAsnProProValProSer-----Pro 376
Db 183 TGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCGTCGATGTTCTNATCAAG 242
Qy 377 LysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 396
Db 243 CGCCTCGCTGGATGATTGAGTCTNTCGATGATTGACCCACCGCAGCAANNACTGGGCTT 302
Qy 397 -TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer---- 414
Db 303 TTACGGACCGGAGATGACAAATTTTGGCGGCCACGATTTCACTGGACACCGGCC 362
Qy 415 -----GlyGlyGlnIleHisIleAlaProMetValLys-GlyLeuGluThrAlaVal 432
Db 363 ACAATGTGTCATCACTCAATTACTTGGCAGCGGATTCGAAGGAATT-----TTTCGT 416
Qy 432 alAsnAspLeuIleSerProProSerGly----- 441
Db 417 TCAAGGACCTGGCCTTTTTCACCGGGGGAAGCGAGCTTTTGGCGCAAGTCGCGAGACG 476
Qy 442 -----AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspV 459
Db 477 ANCAACGGGTTTACTTTTGGCCAGCAAGAACGACCTTGGG----- 517
Qy 459 alThrAlaValProSerThrPheThrSerProValPheThrThrGlyThr-SerVal 478
Db 518 -----ACATTCGCGCAGACGGTTTGGGCAACTCCCACTGGGCCACCTCGACGACGTC 572

Qy 479 AspTyrAla--GlulLeuAsnProSerIleIleValArgAlaGlySer----- 493
Db 573 GACTACGCGCGGAACTTCGTCAGAGCGGTCGTCGCGTTCGCAACACCGCGCGGCAAG 632
Qy 494 -----PheAspProSerSerGlnProAsnAspArgHisValAlaPheS 508
Db 633 CAACAAGGTGGCCATTTTTCGGAACGCGCGCGCGCGCGTNGGAA-----GCAATTC 686
Qy 508 erThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyG 528
Db 687 GAATTAACGCT-----GGMTCCGAACACACCGTTTCCATTGGAACGCGC 728
Qy 528 lyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyG 548
Db 729 GCGCGTGGCTATTTCGCGCGGACGACACGATCTCTGTCGACCGCTCGTCGCGC- 787
Qy 548 lnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProA 568
Db 788 -----GTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCTCCGCTCGAGCTCGCCG 839
Qy 568 laAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnG 588
Db 840 CGGCGCGCTCATGCTCGGCAAGAACACCAAGCGTNTTCTACGCGCGCTCGGAT 899
Qy 588 lyThrPheTyrArgSerThrAspGlyGly 597
Db 900 CGACCTTTTACGTACGCAAGGACACCGCGC 928

RESULT 15

US-10-246-330-3
; Sequence 3, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; TITLE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7407
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7404)
US-10-246-330-3

Alignment Scores:
Pred. No.: 7.2e-12 Length: 7407
Score: 238.00 Matches: 205
Percent Similarity: 32.59% Conservatives: 88
Best Local Similarity: 22.80% Mismatches: 314
Query Match: 5.90% Indels: 292
DB: 16 Gaps: 50

US-09-917-376-3 (1-740) x US-10-246-330-3 (1-7407)

Qy 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30
Db 3250 GTCAACCTGAGCAATGGCAGCAGCTTCAGCGC-----ACCGCGGAGCGCGC 3297
Qy 31 IleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 50
Db 3298 AGCAGCGTGATTTCTCACCAC-----GGCAACGGCAAT 3330
Qy 51 TrpIleProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyr----- 65


```
Qy 627 LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer---ThrAsnGlyGlySerSerTrp 645
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5278 CTCACCGATGGCAATCCGATCGGCCAGACCTTGGCCGACGGCAGCGGCACTGG 5337
Qy 646 SerAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLys 661
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5338 AGCTTCACGCCGGGCACCGCGTGGCCAAACGGCACGGTGGTCAATGCCGTGGCCCGCAGGAC 5397
Qy 662 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr 681
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5398 CCGGCCGGCAATACAGCGGCCCGGCCAGCACCGGTGGACACGGTGGCTCCGGCCACG 5457
Qy 682 GlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHis 701
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5458 -----CCGGTGATCAATCCAGCAAC--- 5478
Qy 702 GlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyr 721
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5479 -----GGCAGCGTGATCACCGGCACC---GCCGAGGTCCGGCCCAAGGTG 5520
Qy 722 IleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5521 ATCCTCACCGATGGCAACGGCAACCCGATCGCGCAGAGCACCGCGCGCAGCGAGTGGT 5577
```

Search completed: October 6, 2005, 04:45:37
Job time : 11542.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:48:48 ; Search time 17,0496 Seconds
(without alignments)
2018.911 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKVKYKNDSPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	99.8	88	7	ADD22925 Acidother
2	469	99.8	89	7	ADD22924 Acidother
3	469	99.8	154	6	ABP73020 Amino aci
4	469	99.8	154	7	ADJ38294 A. cellul
5	469	99.8	762	6	ABP73022 Amino aci
6	469	99.8	762	7	ADJ38291 A. cellul
7	466	99.1	150	6	ABP73025 Amino aci
8	466	99.1	150	8	ADH36640 A. cellul
9	466	99.1	1228	6	ABP73029 Amino aci
10	466	99.1	1228	8	ADH36636 Acidother
11	462	98.3	88	6	ABP73018 Amino aci
12	462	98.3	88	8	ADJ38291 A. cellul
13	462	98.3	89	6	ABP73017 Amino aci
14	462	98.3	89	8	ADJ38291 A. cellul
15	462	98.3	957	7	ABP73015 Amino aci
16	462	98.3	957	7	ADJ38291 A. cellul
17	462	98.3	957	8	ADJ38291 A. cellul
18	436	92.8	153	6	ABP71658 A. cellul
19	436	92.8	153	7	ADJ38291 A. cellul
20	436	92.8	1121	6	ADP75895 Acidother
21	436	92.8	1121	7	ADP75895 Acidother
22	231.5	49.3	782	2	AAR15625 Cellulase
23	195.5	41.6	616	2	AAY13494 Truncated
24	195.5	41.6	616	5	Aae16325 Active ce
25	194.5	41.4	1751	2	AAY13493 Truncated

ALIGNMENTS

RESULT 1
ADD22925
ID ADD22925 standard; protein; 88 AA.

XX AC ADD22925;
XX AC
XX 15-JAN-2004 (first entry)
XX DT
XX DE Acidothermus cellulolyticus avicelase AVIII CBD III #2.
XX enzyme; AVIII; cellulose reduction; agricultural biomass;
KW municipal solid waste; glycoside hydrolase; avicelase.
XX AC
XX Acidothermus cellulolyticus.
XX US2003108988-A1.
XX PN
XX 12-JUN-2003.
XX PD
XX PF 18-OCT-2002; 2002US-00155400.
XX PR
XX 28-JUL-2001; 2001US-00917376.
XX PA (DING/) DING S.
XX PA (ADNE/) ADNEY W S.
XX PA (VINZ/) VINZANT T B.
XX PA (HIMM/) HIMMEL M E.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-810853/76.
XX DR
XX PT New isolated thermal tolerant avicelase polynucleotide useful for
XX PT detection of a polynucleotide encoding AVIII and for reducing cellulose
XX PT in a starting material, e.g. municipal solid waste.
XX PS Claim 16; SEQ ID NO 5; 29pp; English.

The invention relates to an isolated polynucleotide molecule encoding a thermostable AVIII polypeptide. The polynucleotide is useful for detection of a polynucleotide encoding AVIII. The polynucleotide is useful for reducing cellulose in a starting material which involves administering to the starting material, e.g. agricultural biomass or municipal solid waste, a polypeptide molecule of the polynucleotide. The method further comprises administering a second polypeptide molecule chosen from the glycoside hydrolase family of proteins. The present sequence represents the amino acid sequence of Acidothermus cellulolyticus avicelase AVIII CBD III #2.

Aae16324 Active ce
Aay13492 Truncated
Aae16323 Active ce
Aos55907 Bacillus
Ads44721 Bacterial
Aao22444 Protein e
Aao22443 Protein e
Aau98063 Bacillus
Aar42122 NK-1 cell
Aar13227 Novel end
Aaw18790 Corrected
Aar95080 Cellulose
Adg14244 Cellulose
Adg14258 Cellulose
Aay54123 A mammana
Aay28850 Pectate 1
Aay43218 Pectate 1
Aaw5238 Scaffold1
Aaw43108 C. thermo
Adj35112 Xylanase

XX SQ Sequence 88 AA;
 Query Match 99.8%; Score 469; DB 7; Length 88;
 Best Local Similarity 100.0%; Pred. No. 5.4e-46; Indels 0; Gaps 0;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLYNCD 60
 DB 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLYNCD 60
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 RESULT 2
 ADD22924
 ID ADD22924 standard; protein; 89 AA.
 XX AC
 XX ADD22924;
 DT 15-JAN-2004 (first entry)
 XX DE
 DE Acidothermus cellulolyticus avicelase AvIII CBD III #1.
 XX KW
 KW enzyme; AvIII; cellulose reduction; agricultural biomass;
 KW municipal solid waste; glycoside hydrolase; avicelase.
 XX OS
 OS Acidothermus cellulolyticus.
 XX FH
 FH Key Location/Qualifiers
 FT Misc-difference 89
 FT /label= Unknown
 FT
 XX US2003108988-A1.
 XX 12-JUN-2003.
 XX 18-OCT-2002; 2002US-00155400.
 XX 28-JUL-2001; 2001US-00917376.
 XX (DING/) DING S.
 XX (ADNE/) ADNEY W S.
 XX (VINZ/) VINZANT T B.
 XX (HIMM/) HIMMEL M E.
 XX
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;
 XX WPI; 2003-810853/76.
 XX
 XX New isolated thermal tolerant avicelase polynucleotide useful for
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose
 PT in a starting material, e.g. municipal solid waste.
 XX
 XX Claim 16; SEQ ID NO 4; 29pp; English.
 XX
 XX The invention relates to an isolated polynucleotide molecule encoding a
 CC thermostable AvIII polypeptide. The polynucleotide is useful for
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is
 CC useful for reducing cellulose in a starting material which involves
 CC administering to the starting material, e.g. agricultural biomass or
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
 CC method further comprises administering a second polypeptide molecule
 CC chosen from the glycoside hydrolase family of proteins. The present
 CC sequence represents the amino acid sequence of Acidothermus
 CC cellulolyticus avicelase AvIII CBD III #1.
 XX
 XX Sequence 89 AA;
 Query Match 99.8%; Score 469; DB 7; Length 89;
 Best Local Similarity 100.0%; Pred. No. 5.5e-46;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLYNCD 60
 DB 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLYNCD 60
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 RESULT 3
 ABP73020
 ID ABP73020 standard; peptide; 154 AA.
 XX AC
 XX ABP73020;
 DT 03-JUN-2003 (first entry)
 XX DE
 DE Amino acid sequence of the Mana carbohydrate binding domain type III.
 XX KW
 KW Mana; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
 KW food; feed; paper pulp; biofuel; mannanase.
 XX OS
 OS Acidothermus cellulolyticus.
 XX WO2003012110-A1.
 XX PD
 PD 13-FEB-2003.
 XX PF
 PF 28-JUL-2001; 2001WO-US023819.
 XX PR
 PR 28-JUL-2001; 2001WO-US023819.
 XX PA
 PA (MIDE) MIDWEST RES INST.
 XX
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;
 XX WPI; 2003-248182/24.
 XX
 XX Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
 PT cellulolyticus, useful for reducing hemicellulose in a starting material,
 PT for processing of food, and as bulking agents in food stuffs.
 XX
 XX Claim 8; Page 7; 46pp; English.
 XX
 XX The present sequence represents a fragment of Mana, a thermostable
 CC mannanase A polypeptide derived from Acidothermus cellulolyticus. Mana is
 CC a member of the glycoside hydrolase family of enzymes. Mana is useful for
 CC reducing hemicellulose in a starting material to simpler carbohydrate
 CC units, and ultimately to sugars which are useful in the food, feed, paper
 CC pulp, and biofuels industries. It is useful for the processing of food
 CC and in food stuffs as bulking agents, and for the degradation of mannanase.
 CC Mana is also useful to raise polyclonal and monoclonal antibodies that
 CC are useful in purifying Mana, or detecting Mana polypeptide expression,
 CC and as well as reagent tools for characterizing the molecular actions of
 CC Mana polypeptides
 XX
 XX Sequence 154 AA;
 Query Match 99.8%; Score 469; DB 6; Length 154;
 Best Local Similarity 100.0%; Pred. No. 1.1e-45;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLYNCD 60
 DB 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLYNCD 60
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88


```

RESULT 4
ADJ38294
ID ADJ38294 standard; protein; 154 AA.
XX AC
XX ADJ38294;
XX DT
XX 06-MAY-2004 (first entry)
XX DE
XX A. cellulolyticus mannanase ManA carbohydrate binding domain III.
XX KW
XX mannanase A; catalytic domain GH5; carbohydrate binding domain III;
XX KW carbohydrate binding domain II; food processing; foodstuff;
XX KW bulking agent; ManA; enzyme; EC3.2.1.78.
XX XX
XX Acidotherrus cellulolyticus.
XX OS
XX US2003119093-A1.
XX PN
XX 26-JUN-2003.
XX PD
XX 28-JUL-2001; 2001US-00917378.
XX PF
XX 28-JUL-2001; 2001US-00917378.
XX PR
XX (DING/) DING S.
XX PA (ADNE/) ADNEY W S.
XX PA (VINZ/) VINZANT T B.
XX PA (HIMM/) HIMMEL M E.
XX XX
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
XX PI WPI; 2003-863435/80.
XX DR
XX New composition comprising a purified mannanase A peptide, comprising a
XX PT catalytic domain GH5, carbohydrate binding domain III or carbohydrate
XX PT binding domain II, useful in food processing and in food stuffs as
XX PT bulking agents.
XX XX
XX Claim 7; SEQ ID NO 4; 18pp; English.
XX PS
XX The invention relates to a composition comprising a purified mannanase A
XX CC peptide, having a catalytic domain GH5, carbohydrate binding domain III
XX CC or carbohydrate binding domain II. The composition is useful in food
XX CC processing and in foodstuffs as bulking agents. The present sequence
XX CC represents the amino acid sequence of A. cellulolyticus mannanase ManA
XX CC carbohydrate binding domain III.
XX XX
XX Sequence 154 AA;
XX SQ
Query Match 99.8%; Score 469; DB 7; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.1e-45;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 60
Db 1 VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 60
QY 61 WAAMGCGNIRASFGSVNPTATPTADTYLQ 88
Db 61 WAAMGCGNIRASFGSVNPTATPTADTYLQ 88
RESULT 5
ABP73022
ID ABP73022 standard; protein; 762 AA.
XX AC
XX ABP73022;
XX DT
XX 03-JUN-2003 (first entry)
XX XX
XX Amino acid sequence of the ManA polypeptide.
XX DE
XX ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
KW

```

```

KW food; feed; paper pulp; biofuel; mannanase.
XX Acidotherrus cellulolyticus.
XX OS
XX WO2003012110-A1.
XX PN
XX 13-FEB-2003.
XX PD
XX 28-JUL-2001; 2001WO-US023819.
XX PF
XX 28-JUL-2001; 2001WO-US023819.
XX PR
XX (MIDE ) MIDWEST RES INST.
XX PA
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
XX PI WPI; 2003-248182/24.
XX DR N-PSDB; ABZ77633.
XX DR
XX Novel thermal tolerant mannanase A polypeptide derived from Acidotherrus
XX PT cellulolyticus, useful for reducing hemicellulose in a starting material,
XX PT for processing of food, and as bulking agents in food stuffs.
XX XX
XX Claim 3; Page 18-19; 46pp; English.
XX PS
XX The present sequence represents ManA, a thermostable mannanase A
XX CC polypeptide derived from Acidotherrus cellulolyticus. ManA is a member of
XX CC the glycoside hydrolase family of enzymes. ManA is useful for reducing
XX CC hemicellulose in a starting material to simpler carbohydrate units, and
XX CC ultimately to sugars which are useful in the food, feed, paper pulp, and
XX CC biofuels industries. It is useful for the processing of food and in food
XX CC stuffs as bulking agents, and for the degradation of mannanase. ManA is
XX CC also useful to raise polyclonal and monoclonal antibodies that are useful
XX CC in purifying ManA, or detecting ManA polypeptide expression, and as well
XX CC as reagent tools for characterizing the molecular actions of ManA
XX CC polypeptides
XX XX
XX Sequence 762 AA;
XX SQ
Query Match 99.8%; Score 469; DB 6; Length 762;
Best Local Similarity 100.0%; Pred. No. 7.8e-45;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 60
Db 455 VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSTLVNCD 514
QY 61 WAAMGCGNIRASFGSVNPTATPTADTYLQ 88
Db 515 WAAMGCGNIRASFGSVNPTATPTADTYLQ 542
RESULT 6
ADJ38291
ID ADJ38291 standard; protein; 762 AA.
XX AC
XX ADJ38291;
XX DT
XX 06-MAY-2004 (first entry)
XX XX
XX A. cellulolyticus mannanase ManA.
XX DE
XX mannanase A; catalytic domain GH5; carbohydrate binding domain III;
XX KW carbohydrate binding domain II; food processing; foodstuff;
XX KW bulking agent; ManA; enzyme; EC3.2.1.78.
XX XX
XX Acidotherrus cellulolyticus.
XX OS
XX US2003119093-A1.
XX PN
XX 26-JUN-2003.
XX PD
XX 28-JUL-2001; 2001US-00917378.
XX PF

```


XX 28-JUL-2001; 2001US-00917378.
 XX (DING/) DING S.
 XX (ADNE/) ADNEY W S.
 XX (VINZ/) VINZANT T B.
 XX (HIMM/) HIMMEL M E.
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;
 XX WPI; 2003-863435/80.
 XX N-PSDB; ADJ38292.
 XX New composition comprising a purified mannanase A peptide, comprising a
 XX catalytic domain GH5, carbohydrate binding domain III or carbohydrate
 XX binding domain II, useful in food processing and in food stuffs as
 XX bulking agents.
 XX Claim 10; SEQ ID NO 1; 18pp; English.
 XX The invention relates to a composition comprising a purified mannanase A
 XX peptide, having a catalytic domain GH5, carbohydrate binding domain III
 XX or carbohydrate binding domain II. The composition is useful in food
 XX processing and in foodstuffs as bulking agents. The present sequence
 XX represents the amino acid sequence of A. cellulolyticus mannanase Mana.
 XX Sequence 762 AA;
 SQ
 Query Match 99.8%; Score 469; DB 7; Length 762;
 Best Local Similarity 100.0%; Pred. No. 7.8e-45;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VSGGVKQVQKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLYVNC 60
 DB 455 VSGGVKQVQKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLYVNC 514
 QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
 DB 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542
 RESULT 7
 ABP73025
 ID ABP73025 standard; peptide; 150 AA.
 XX ABP73025;
 XX 03-JUN-2003 (first entry)
 XX Amino acid sequence of the GuxA carbohydrate binding domain type III.
 XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
 XX detergent; pulp processing; paper processing; feed processing; textile.
 XX Acidothermus cellulolyticus.
 XX WO2003012109-A1.
 XX 13-FEB-2003.
 XX 28-JUL-2001; 2001WO-US023817.
 XX 28-JUL-2001; 2001WO-US023817.
 XX (MIDE) MIDWEST RES INST.
 XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
 XX WPI; 2003-239526/23.
 XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus
 XX cellulolyticus, useful for reducing cellulose in a starting material, and
 XX for the conversion of biomass to biofuels and biofuel additives.

XX Claim 6; Page 7; 47pp; English.
 XX ABP73023-28 represent peptide fragments of a GuxA polypeptide. GuxA is a
 XX thermostable cellulase, and is a member of the glycoside hydrolase family
 XX of enzymes. GuxA is useful for reducing cellulose in a starting material
 XX such as agricultural biomass to sugars. This is useful in biofuel
 XX production. GuxA is also useful in the conversion of biomass to biofuels
 XX and biofuel additives, in detergents, pulp and paper processing, food and
 XX feed processing, and in textile process. GuxA is also useful for raising
 XX polyclonal and monoclonal antibodies that are useful in purifying GuxA,
 XX or detecting GuxA polypeptide expression, as well as reagent tools for
 XX characterizing the molecular actions of GuxA polypeptides
 SQ Sequence 150 AA;
 Query Match 99.1%; Score 466; DB 6; Length 150;
 Best Local Similarity 98.9%; Pred. No. 2.3e-45;
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VSGGVKQVQKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLYVNC 60
 DB 1 VSGGVKQVQKNDSPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLYVNC 60
 QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
 DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
 RESULT 8
 ADH36640
 ID ADH36640 standard; protein; 150 AA.
 XX ADH36640;
 XX 11-MAR-2004 (first entry)
 XX A. cellulolyticus glycoside hydrolase, GuxA CBD type III domain.
 XX Thermal tolerant cellulase; glycoside hydrolase; GuxA;
 XX carbohydrate degradation; cellulose; agricultural biomass;
 XX municipal solid waste; thermostable; enzyme;
 XX carbohydrate binding domain type III; CBD III.
 XX Acidothermus cellulolyticus.
 XX US2003104522-A1.
 XX 05-JUN-2003.
 XX 28-JUL-2001; 2001US-00917383.
 XX 28-JUL-2001; 2001US-00917383.
 XX (DING/) DING S.
 XX (ADNE/) ADNEY W S.
 XX (VINZ/) VINZANT T B.
 XX (HIMM/) HIMMEL M E.
 XX (DECK/) DECKER S R.
 XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
 XX WPI; 2004-106451/11.
 XX Novel isolated thermostable GuxA polypeptide useful for detecting
 XX polynucleotide encoding GuxA, assessing carbohydrate degradation activity
 XX of GuxA, reducing cellulose in starting material e.g., agricultural
 XX biomass.
 XX Claim 27; SEQ ID NO 5; 20pp; English.
 XX The present invention relates to the isolation of a thermal tolerant
 XX cellulase from Acidothermus cellulolyticus. The cellulase is a member of

CC the glycoside hydrolase family and is designated GuxA. Also disclosed are
 CC methods of producing and using GuxA. The GuxA polypeptide is useful for
 CC the detection of a polynucleotide encoding GuxA. The polypeptide sequence
 CC is also useful for assessing the carbohydrate, e.g. cellulose,
 CC degradation activity of GuxA. GuxA is useful for reducing cellulose in a
 CC starting material such as agricultural biomass or municipal solid waste.
 CC The polypeptide molecule of GuxA is thermostable. The present sequence
 CC represents a domain of the A. cellulolyticus GuxA polypeptide.
 XX
 SQ Sequence 150 AA;

Query Match 99.1%; Score 466; DB 8; Length 150;
 Best Local Similarity 98.9%; Pred. No. 2.3e-45;
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
 DB 1 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
 QY 61 WAAMGCGNIRASFGSVNPTATPTADTYLQ 88
 DB 61 WAAMGCGNIRASFGSVNPTATPTADTYLQ 88

RESULT 9
 ABP73029
 ID ABP73029 standard; protein; 1228 AA.

AC ABP73029;

DT 03-JUN-2003 (first entry)

XX Amino acid sequence of the GuxA polypeptide.

XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
 KW detergent; pulp processing; paper processing; feed processing; textile.

XX Acidothermus cellulolyticus.

XX W02003012109-A1.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023817.

XX 28-JUL-2001; 2001WO-US023817.

XX (MIDE) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2003-239526/23.

XX N-PSDB; ABZ77634.

XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus
 PT cellulolyticus, useful for reducing cellulose in a starting material, and
 PT for the conversion of biomass to biofuels and biofuel additives.

PS Claim 3; Page 19; 47pp; English.

XX The present sequence represents a GuxA polypeptide. GuxA is thermostable
 CC cellulase, and is a member of the glycoside hydrolase family of enzymes.
 CC GuxA is useful for reducing cellulose in a starting material such as
 CC agricultural biomass to sugars. This is useful in biofuel production.
 CC GuxA is also useful in the conversion of biomass to biofuels and biofuel
 CC additives, in detergents, pulp and paper processing, food and feed
 CC processing, and in textile process. GuxA is also useful for raising
 CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,
 CC or detecting GuxA polypeptide expression, as well as reagent tools for
 CC characterizing the molecular actions of GuxA polypeptides

XX Sequence 1228 AA;

Query Match 99.1%; Score 466; DB 6; Length 1228;
 Best Local Similarity 98.9%; Pred. No. 3.1e-44;
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60
 DB 584 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 643
 QY 61 WAAMGCGNIRASFGSVNPTATPTADTYLQ 88
 DB 644 WAAMGCGNIRASFGSVNPTATPTADTYLQ 671

RESULT 10
 ADH36636

ID ADH36636 standard; protein; 1228 AA.

XX ADH36636;

XX 11-MAR-2004 (first entry)

XX Acidothermus cellulolyticus glycoside hydrolase, GuxA.

XX Thermal tolerant cellulase; glycoside hydrolase; GuxA;
 KW carbohydrate degradation; cellulose; agricultural biomass;
 KW municipal solid waste; thermostable; enzyme.

XX Acidothermus cellulolyticus.

XX US2003104522-A1.

XX 05-JUN-2003.

XX 28-JUL-2001; 2001US-00917383.

XX 28-JUL-2001; 2001US-00917383.

XX (DING/) DING S.

XX (ADNE/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX (DECK/) DECKER S R.

XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2004-106451/11.

XX N-PSDB; ADH36637.

XX Novel isolated thermostable GuxA polypeptide useful for detecting
 PT polynucleotide encoding GuxA, assessing carbohydrate degradation activity
 PT of GuxA, reducing cellulose in starting material e.g., agricultural
 PT biomass.

XX Claim 27; SEQ ID NO 1; 20pp; English.

XX The present invention relates to the isolation of a thermal tolerant
 CC cellulase from Acidothermus cellulolyticus. The cellulase is a member of
 CC the glycoside hydrolase family and is designated GuxA. Also disclosed are
 CC methods of producing and using GuxA. The GuxA polypeptide is useful for
 CC the detection of a polynucleotide encoding GuxA. The polypeptide sequence
 CC is also useful for assessing the carbohydrate, e.g. cellulose,
 CC degradation activity of GuxA. GuxA is useful for reducing cellulose in a
 CC starting material such as agricultural biomass or municipal solid waste.
 CC The polypeptide molecule of GuxA is thermostable. The present sequence
 CC represents A. cellulolyticus glycoside hydrolase, GuxA.

XX Sequence 1228 AA;

Query Match 99.1%; Score 466; DB 8; Length 1228;
 Best Local Similarity 98.9%; Pred. No. 3.1e-44;
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSTLVYVNC 60

Db 584 VSGGLKQVYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 643
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671

RESULT 11
 ABP73018
 ID ABP73018 standard; peptide; 88 AA.
 XX AC ABP73018;
 XX DT 03-JUN-2003 (first entry)
 XX DE Amino acid sequence of AvIII carbohydrate binding domain type II.
 XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
 KW detergent; pulp processing; paper processing; feed processing; textile;
 KW cellulose.
 XX OS Acidothermus cellulolyticus.
 XX PN WO2003012090-A2.
 XX PD 13-FEB-2003.
 XX PF 28-JUL-2001; 2001WO-US023818.
 XX PR 28-JUL-2001; 2001WO-US023818.
 XX PA (MIDE) MIDWEST RES INST.
 XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
 XX WPI; 2003-248177/24.

XX New thermostable AvIII peptide from Acidothermus cellulolyticus, useful
 PT for degradation of cellulose or in generating anti-AvIII antibodies for
 PT purifying recombinant AvIII polypeptides from genetically engineered
 PT host cells.
 XX Claim 6; Page 8; 4pp; English.
 XX The present sequence is derived from a thermostable avicelase, designated
 CC AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,
 CC and is a cellulase. AvIII is useful in the conversion of biomass to
 CC biofuels and biofuel additives. It may be useful in the production of
 CC detergents, pulp and paper processing, food and feed processing and in
 CC textile processes. The thermostable AvIII peptide is useful in the
 CC degradation of cellulose, and in generating specific anti-AvIII
 CC antibodies that are useful in purifying recombinant AvIII polypeptides
 CC from genetically engineered host cells, in detecting AvIII polypeptide
 CC expression, as well as a reagent tool for characterizing the molecular
 CC actions of the polypeptide. The AvIII polynucleotide is useful as a
 CC source of probes or primers in various diagnostic assays
 XX Sequence 88 AA;

Query Match 98.3%; Score 462; DB 6; Length 88;
 Best Local Similarity 97.7%; Pred. No. 3.5e-45;
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VSGGVKQVYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60
 Db 1 VSGGVKQVYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 Db 61 WAAICGNIRASFGSVNPATPTADTYLQ 88

RESULT 12
 ADOS2317
 ID ADOS2317 standard; protein; 88 AA.
 XX AC ADOS2317;
 XX DT 15-JUL-2004 (first entry)
 XX DE A. cellulolyticus AvIII CBD_III partial protein #2.
 XX KW Thermostable cellulase; avicelase III; AvIII; cellulose reduction;
 KW biomass degradation; ethanol formation; industrial chemical;
 KW fabric treatment; carbohydrate binding domain type III; CBD III.
 XX OS Acidothermus cellulolyticus.
 XX PN US2004038334-A1.
 XX PD 26-FEB-2004.
 XX PF 28-JUL-2001; 2001US-00917376.
 XX PR 28-JUL-2001; 2001US-00917376.
 XX PA (DING/) DING S.
 PA (ADNEY/) ADNEY W S.
 PA (VINZ/) VINZANT T B.
 PA (HIMM/) HIMMEL M E.
 XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
 XX WPI; 2004-203224/19.
 XX Novel thermostable AvIII polypeptide of glycoside hydrolase family and
 PT isolated from Acidothermus cellulolyticus, useful for degrading treated
 PT biomass into simpler forms of carbohydrate.
 XX Claim 28; SEQ ID NO 5; 19pp; English.

XX The invention relates to a thermostable cellulase enzyme, avicelase III
 CC (AvIII) and its nucleic acid sequence. AvIII is useful for reducing
 CC cellulose in a starting material. A thermostable AvIII peptide is useful
 CC for degrading treated biomass into simpler forms of carbohydrate, which
 CC is used in the formation of ethanol or other industrial chemicals. It is
 CC also useful for treating fabrics to remove cellulose-containing stains.
 CC The present sequence is Acidothermus cellulolyticus AvIII carbohydrate
 CC binding domain type III (CBD_III) partial protein.
 XX Sequence 88 AA;

Query Match 98.3%; Score 462; DB 8; Length 88;
 Best Local Similarity 97.7%; Pred. No. 3.5e-45;
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VSGGVKQVYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60
 Db 1 VSGGVKQVYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 Db 61 WAAICGNIRASFGSVNPATPTADTYLQ 88

RESULT 13
 ABP73017
 ID ABP73017 standard; peptide; 89 AA.
 XX AC ABP73017;
 XX DT 03-JUN-2003 (first entry)
 XX DE Amino acid sequence of AvIII carbohydrate binding domain type III.
 XX

KW Avicelase; AviiiI; glycoside hydrolase; enzyme; cellulase; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
XX cellulose.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

FT Misc-difference 89 /note= "unspecified residue"

XX WO2003012090-A2.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023818.

XX 28-JUL-2001; 2001WO-US023818.

XX (MIDE) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-248177/24.

XX New thermostable AviiiI peptide from Acidothermus cellulolyticus, useful
PT for degradation of cellulose or in generating anti-AviiiI antibodies for
PT purifying recombinant AviiiI polypeptides from genetically engineered
PT host cells.

XX Claim 6; Page 8; 4pp; English.

XX The present sequence is derived from a thermostable avicelase, designated
CC AviiiI. AviiiI is a member of the glycoside hydrolase family of enzymes,
CC and is a cellulase. AviiiI is useful in the conversion of biomass to
CC biofuels and biofuel additives. It may be useful in the production of
CC detergents, pulp and paper processing, food and feed processing and in
CC textile processes. The thermostable AviiiI peptide is useful in the
CC degradation of cellulose, and in generating specific anti-AviiiI
CC antibodies that are useful in purifying recombinant AviiiI polypeptides
CC from genetically engineered host cells, in detecting AviiiI polypeptide
CC expression, as well as a reagent tool for characterizing the molecular
CC actions of the polypeptide. The AviiiI polynucleotide is useful as a
CC source of probes or primers in various diagnostic assays

XX Sequence 89 AA;

Query Match 98.3%; Score 462; DB 6; Length 89;
Best Local Similarity 97.7%; Pred. No. 3.5e-45;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGGSSTLVYVNC 60

Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGGSSTLVYVNC 60

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 14

AD052316

ID AD052316 standard; protein; 89 AA.

XX AC AD052316;

XX 15-JUL-2004 (first entry)

XX A. cellulolyticus AviiiI CBD_III partial protein #1.

XX Thermostable cellulase; avicelase III; AviiiI; cellulose reduction;
KW biomass degradation; ethanol formation; industrial chemical;
KW fabric treatment; carbohydrate binding domain type III; CBD III.

OS Acidothermus cellulolyticus.

XX Key Location/Qualifiers

FT Misc-difference 89 /note= "A string of unknown amino acids (50 or more)"

XX US2004038334-A1.

XX 26-FEB-2004.

XX 28-JUL-2001; 2001US-00917376.

XX 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

XX (ADNE/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2004-203224/19.

XX Novel thermostable AviiiI polypeptide of glycoside hydrolase family and
PT isolated from Acidothermus cellulolyticus, useful for degrading treated
PT biomass into simpler forms of carbohydrate.

XX Claim 28; SEQ ID NO 4; 19pp; English.

XX The invention relates to a thermostable cellulase enzyme, avicelase III
CC (AviiiI) and its nucleic acid sequence. AviiiI is useful for reducing
CC cellulose in a starting material. A thermostable AviiiI peptide is useful
CC for degrading treated biomass into simpler forms of carbohydrate, which
CC is used in the formation of ethanol or other industrial chemicals. It is
CC also useful for treating fabrics to remove cellulose-containing stains.
CC The present sequence is Acidothermus cellulolyticus AviiiI carbohydrate
CC binding domain type III (CBD_III) partial protein.

XX Sequence 89 AA;

Query Match 98.3%; Score 462; DB 8; Length 89;

Best Local Similarity 97.7%; Pred. No. 3.5e-45;

Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGGSSTLVYVNC 60

Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGGSSTLVYVNC 60

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 15

ABP73015

ID ABP73015 standard; protein; 957 AA.

XX AC ABP73015;

XX 03-JUN-2003 (first entry)

XX Amino acid sequence of the avicelase AviiiI.

XX Avicelase; AviiiI; glycoside hydrolase; enzyme; cellulase; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
KW cellulose.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

FT Misc-difference 957

FT /note= "unspecified residue encoded by N"

XX

PN WO2003012090-A2.
XX 13-FEB-2003.
XX 28-JUL-2001; 2001WO-US023818.
PF XX
XX 28-JUL-2001; 2001WO-US023818.
PR XX
XX (WIDE) MIDWEST RES INST.
PA
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
XX WPI; 2003-248177/24.
DR N-PSDB; ABZ77632.
DR XX
XX New thermostable Aviii peptide from Acidothermus cellulolyticus, useful
PT for degradation of cellulose or in generating anti-Aviii antibodies for
PT purifying recombinant Aviii polypeptides from genetically engineered
PT host cells.
XX
XX Claim 2; Page 20; 44pp; English.
PS
CC The present sequence represents a thermostable avicelase polypeptide,
CC designated Aviii. Aviii is a member of the glycoside hydrolase family
CC of enzymes, and is a cellulase. Aviii is useful in the conversion of
CC biomass to biofuels and biofuel additives. It may be useful in the
CC production of detergents, pulp and paper processing, food and feed
CC processing and in textile processes. The thermostable Aviii peptide is
CC useful in the degradation of cellulose, and in generating specific anti-
CC Aviii antibodies that are useful in purifying recombinant Aviii
CC polypeptides from genetically engineered host cells, in detecting Aviii
CC polypeptide expression, as well as a reagent tool for characterizing the
CC molecular actions of the polypeptide. The Aviii polynucleotide is useful
CC as a source of probes or primers in various diagnostic assays
XX
SQ Sequence 957 AA;
Query Match 98.3%; Score 462; DB 6; Length 957;
Best Local Similarity 97.7%; Pred. No. 6.6e-44;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
Db 869 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 928
Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956
Search completed: October 5, 2005, 08:03:11
Job time : 18.0496 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	191	40.6	1711	2	T31337	1,4-beta-glucanase
2	190.5	40.5	1779	2	T31085	xylanase - Caldiche
3	190	40.4	508	2	G69593	cellulase (EC 3.2.
4	190	40.4	508	2	A28874	cellulase (EC 3.2.
5	188	40.0	499	2	JN0111	cellulase (EC 3.2.
6	186.5	39.7	1331	2	A48954	mannan endo-1,4-be
7	186.5	39.7	1742	2	T17120	cellulase (EC 3.2.
8	186	39.6	499	2	A27198	cellulase (EC 3.2.
9	182.5	38.8	915	2	A43802	cellulase (EC 3.2.
10	182.5	38.8	1039	2	S02711	cellulase (EC 3.2.
11	180.5	38.4	145	2	A41897	cellulase homolog
12	180	38.3	486	2	I40548	bifunctional cellul
13	172	36.6	700	2	B41897	cellulase (EC 3.2.
14	165.5	35.2	879	2	A47704	endoglucanase I (E
15	165	35.1	1854	2	S36859	c1pA protein - Clo
16	149.5	31.8	505	2	S39962	endoglucanase - Er
17	147.5	31.4	504	2	S54744	cellulase (EC 3.2.
18	141.5	30.1	986	2	S12021	thermoactive cellul
19	130.5	27.8	586	2	PC6006	scaffolding protei
20	130.5	27.8	1483	2	C97012	probably celluloso
21	126.5	26.9	1162	2	T30433	scaffolding protei
22	117	24.9	1230	2	S47466	cellulose 1,4-beta
23	114	24.3	1848	2	A44140	cellulose-binding
24	75.5	16.1	618	2	T08685	hypothetical prote
25	73.5	15.6	547	2	T25478	hypothetical prote
26	73.5	15.6	1428	2	AC2324	probable RTX fami
27	70	14.9	5188	2	B85547	hypothetical prote
28	70	14.9	5291	2	F90696	cellulurface glyco
29	68.5	14.6	574	2	A69196	

Best Local Similarity 43.5%; Pred. No. 1.4e-11;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGQLVNTGSSVDLSVTTVRYWFFTRDGGSSTLVYNCDDWA 63
1113 GLKLYKNNETASTSRPFWKTVNGGSSVDLSRVKIRYWTVDGDKPOSAY-CDWAQ 1171

Db 1113 GLKLYKNNETASTSRPFWKTVNGGSSVDLSRVKIRYWTVDGDKPOSAY-CDWAQ 1171

QY 64 MGCNIRASFGSVNPNATPTADTYLQ 88
1172 IGASNVTFNFKLITSGVSGADYYLE 1196

Db 1172 IGASNVTFNFKLITSGVSGADYYLE 1196

RESULT 3

G69593

Cellulase (EC 3.2.1.4) bgIC precursor - Bacillus subtilis

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004

C:Accession: G69593; A26114; I40353; S24239; S49103; I39803

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteaux, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choudhury, M.; Emmerison, P.T.; Enian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallet, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schoeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Seron, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: G69593

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-508 <KUN>

A:Cross-references: UNIPROT:P10475; UNIPROT:Q03882; GB:Z99113; GB:AL009126; NID:G2634090

A:Experimental source: strain 168

R:MacKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moranelli, F.; Seligman, A.; Nucleic Acids Res. 14, 9159-9170, 1986

A:Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.

A:Reference number: A26114; MUID:87066783; PMID:3024130

A:Accession: A26114

A:Molecule type: DNA

A:Residues: 10-508 <MAC>

A:Cross-references: GB:X04689; NID:G39823; PIDN:CAA28392.1; PID:G39824

A:Experimental source: strain PAP115

A:Note: part of this sequence, including the amino end of the mature form, was confirmed by Lindahl, V.; Aa, K.; Tronemo, A.

A:Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis C

A:Reference number: I40353; MUID:95225656; PMID:7710280

A:Accession: I40353

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 10-291, 'N', 293-508 <LIN1>

A:Cross-references: EMBL:X67044; NID:G39776; PIDN:CAA47429.1; PID:G39777

R:Lindahl, V.; Aa, K.

submitted to the EMBL Data Library, June 1992

A:Reference number: S24239

A:Accession: S24239

A:Molecule type: DNA

A:Residues: 10-291, 'N', 293-508 <LIN2>

A:Cross-references: EMBL:X67044; NID:G39776; PIDN:CAA47429.1; PID:G39777

R:Wolf, M.; Geczi, A.; Borris, R.

A:Experimental source: strain CK-2

A:Submitted to the EMBL Data Library, December 1993

A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons

A:Reference number: S49103

A:Accession: S49103

A:Molecule type: DNA

A:Residues: 10-508 <WOL>

A:Cross-references: EMBL:Z29076; NID:G509266; PIDN:CAA82317.1; PID:G509267

R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.

Korean J. Microbiol. 24, 236-242, 1986

A:Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis

A:Reference number: I39803

A:Accession: I39803

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'F', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ', 36

A:Cross-references: GB:M38634; NID:G142657; PIDN:AAA2300.1; PID:G142658

A:Experimental source: strain ATCC 6633

C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal processing

C:Genetics: bgIC

A:Gene: bgIC

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose

A:Pathway: cellulose degradation

C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation

F:1-38/Domain: signal sequence #status predicted <SIG>

F:39-508/Product: cellulase #status predicted <MAT>

Query Match 40.4%; Score 190; DB 2; Length 508;
Best Local Similarity 40.7%; Pred. No. 4.2e-12;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSPAGDNQIKPGQLVNTGSSVDLSVTTVRYWFFTRDGGSSTLVYNCDDWA 62
362 GISVQYRAGDSMNSNQIRPQLIKNNGNATVDLKDVTARYWNVKNKGN---FDCDYA 418

Db 362 GISVQYRAGDSMNSNQIRPQLIKNNGNATVDLKDVTARYWNVKNKGN---FDCDYA 418

QY 63 AMGCNIRASFGSVNPNATPTADTYLQ 88
419 QIGCGNVTHKFTVTLHKPKQADTYLE 444

Db 419 QIGCGNVTHKFTVTLHKPKQADTYLE 444

RESULT 4

A26874

cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Bacillus subtilis

C>Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 09-Jul-2004

C:Accession: A26874; B26874

R:Robson, L.M.; Chambliss, G.H.

J. Bacteriol. 169, 2017-2025, 1987

A:Title: Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.

A:Reference number: A26874; MUID:87194581; PMID:3106328

A:Accession: A26874

A:Molecule type: DNA

A:Residues: 1-508 <ROB1>

A:Cross-references: UNIPROT:P07983; GB:M16185; NID:G143007; PIDN:AAA22496.1; PID:G143008

A:Experimental source: strain DLG

A:Accession: B26874

A:Molecule type: protein

A:Residues: 39-53 <ROB2>

A:Experimental source: strain DLG

A:Note: the authors believe Met-1 and Met-2 may be alternate initiators

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose

A:Pathway: cellulose degradation

C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; polysaccharide degradation

F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 40.4%; Score 190; DB 2; Length 508;
Best Local Similarity 43.0%; Pred. No. 4.2e-12;
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSPAGDNQIKPGQLVNTGSSVDLSVTTVRYWFFTRDGGSSTLVYNCDDWA 62
362 GISVQYKAGDGRVNSNQIRPQLIKNNGNATVDLKDVTARYWNVKNKGN---FDCDYA 418

Db 362 GISVQYKAGDGRVNSNQIRPQLIKNNGNATVDLKDVTARYWNVKNKGN---FDCDYA 418

QY 63 AMGCNIRASFGSVNPNATPTADTYLQ 88

Query Match 39.6%; Score 186; DB 2; Length 499;
Best Local Similarity 40.7%; Pred. NO. 1.le-11;
Matches 35; Conservative 19; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTWRYWF-TRDGGSSLVYNCDWA 62
Db : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
353 GISVQYKAGDGGVNSQIRPOLHIKNNGHATVDLKDTIARYWYNAKKQN---FDCDYA 409

QY 63 AMCGGNIRASFGSVNPATPTADTYLQ 88
Db : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
410 QIGCGNLTHKFVTLHKPKQGADTYLE 435

RESULT 9
A43802
Cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Caldocellum saccharolyticum
C;Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C;Accession: A43802
R;Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 56, 3117-3124, 1990
A;Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile
A;Reference number: A43802; MUID:91136262; PMID:2126700
A;Accession: A43802
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: DNA
A;Residues: 1-915 <SAU>
A;Cross-references: EMBL:X13602
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A;Pathway: cellulose degradation
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 38.8%; Score 182.5; DB 2; Length 915;
Best Local Similarity 41.9%; Pred. NO. 4.7e-11;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 3 GGKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTWRYWFTRDGGSSLVYNCDWA 62
Db : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
349 GOIKVLANKETNSTTIRPLKVVNSSSSIDLSTRTIARYWTVDGERAQSAVS-DWA 407

QY 63 AMCGGNIRASFGSVNPATPTADTYLQ 88
Db : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
408 QIGASNVTPEFKLSLVSSVSGADTYLE 433

RESULT 10
S02711
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
N;Alternate names: endo-1,4-beta-glucanase
N;Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)
C;Species: Caldocellum saccharolyticum
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C;Accession: S02711
R;Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
Nucleic Acids Res. 17, 439, 1989
A;Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo
A;Reference number: S02711; MUID:89098398; PMID:2789517
A;Accession: S02711
A;Molecule type: DNA
A;Residues: 1-1039 <SAU>
A;Cross-references: UNIPROT:P10474; EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646
C;Genetics:
A;Gene: celB
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A;Pathway: cellulose degradation
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1039/Product: cellulase #status predicted <MAR>
F;72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

```

Query Match          39.8%, Score 182.5, DB 2, Length 1039;
Best Local Similarity 41.9%; Pred. No. 5.4e-11;
Matches 36; Conservative 18; Mismatches 31; Indels 1; Gaps 1
QY      3 GGKVKVQYKNNDSPAGDNOIKPGLQLVNTGSSVDLSITTVRYWFTEDGGSSSTLIVNCDWA 62
Db       | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
         420 GQIKVLVANKETNSTTTIRPLWKVNVSGSSSIDUSRTVIRYWTVDGERAQSASVS-DWA 478

QY      63 AMGCCGNIRASFSGSVNPATPTADTYLQ 88
Db       | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
         479 QIGASNVTFFKVLKSSSVSGADYYLE 504

RESULT 11
A41897
cellulase homolog - Bacillus lautus (fragment)
C:Species: Bacillus lautus
C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jul-2004
C:Accession: A41897; S27498
R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A>Title: cglA from Bacillus lautus PL236 encodes a novel cellulose-binding endo
A:Reference number: A41897; PMID:92276330; PMID:1592807
A:Accession: A41897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <HAN>
A:Cross-references: UNIPROT:P29718; EMBL:M76588; NID:g142661; PIDN:AAA22302.1;
A:Experimental source: PL236
A>Note: sequence extracted from NCBI backbone (NCBIP:104604)
```

[illegible]

Db 411 LGCNVSHVTVLHKPKQADTYLE 435

RESULT 13

B41897

cellulase (EC 3.2.1.4) - Bacillus lautus

N;Alternate names: endo-1,4-beta-glucanase

C;Species: Bacillus lautus

C;Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 09-Jul-2004

C;Accession: B41897; S27499

R;Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.

J. Bacteriol. 174, 3522-3531, 1992

A;Title: celsa from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4

A;Reference number: A41897; MUID:92276330; PMID:1592807

A;Accession: B41897

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-700 <HAN>

A;Cross-references: UNIPROT:P29719; EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142661

A;Experimental source: PL236

A;Note: sequence extracted from NCBI backbone (NCBIP:104605)

C;Function:

A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel

A;Pathway: cellulose degradation

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.6%; Score 172; DB 2; Length 700;

Best Local Similarity 44.9%; Pred. No. 4.4e-10;

Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVYNC 59

Db 549 VNSDLVQYKGDNRNATDQIKPHFNQKSPVLSLTLRYFTKD-SSAAMNGWI 607

QY 60 DWAMGCGNIRASFGSVNPATPTADTYLQ 88

Db 608 DWAKLGSGNIQISFGNHA--DSDTYAB 634

RESULT 14

A47704

endoglucanase I (EC 3.2.1.-) Cell - Clostridium thermoceum

C;Species: Clostridium thermoceum

C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C;Accession: A47704

R;Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Huskisson, N.S.; Gilbert, H.J.

J. Gen. Microbiol. 139, 307-316, 1993

A;Title: Gene sequence and properties of Cell, a family E endoglucanase from Clostridium

A;Reference number: A47704; MUID:93171873; PMID:8436949

A;Accession: A47704

A;Status: preliminary

A;Molecule type: DNA; protein

A;Residues: 1-879 <HAZ>

A;Cross-references: UNIPROT:Q02934; GB:I04735; NID:g144807; PIDN:AAA20892.1; PID:g144808

A;Note: sequence extracted from NCBI backbone (NCBIN:125637; NCBIP:125638)

C;Keywords: glycosidase; hydrolase

Query Match 35.2%; Score 165.5; DB 2; Length 879;

Best Local Similarity 33.0%; Pred. No. 2.7e-09;

Matches 29; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 1 VSGGVKVOYKNDSPGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVYNC 60

Db 736 IKGEVLQYANGAGTNSINFRKLIINNGKAINLSVDKIRYYTYKGGASQNF-CD 794

QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88

Db 795 WSSAGNSNVTGNFFNLSPKEGADTCLE 822

RESULT 15

S36859

cipA protein - Clostridium thermoceum

N;Alternate names: probable cellulosome protein large chain SL

C;Species: Clostridium thermoceum

C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004

C;Accession: S36859; S33527; S25767; S28659; T18261

R;Gerngross, U.T.; Demain, A.L.

submitted to the EMBL Data Library, January 1993

A;Reference number: S36859

A;Accession: S36859

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1854 <GER>

A;Cross-references: UNIPROT:Q06851; EMBL:L08665

R;Gerngross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.

Mol. Microbiol. 8, 325-334, 1993

A;Title: Sequencing of a Clostridium thermoceum gene (cipA) encoding the cellulosomal

A;Reference number: S33527; MUID:93302508; PMID:8316083

A;Accession: S33527

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1615,1617-1854 <GE2>

A;Cross-references: EMBL:L08665

R;Fujino, T.; Beguin, P.; Aubert, J.P.

FEMS Microbiol. Lett. 94, 165-170, 1992

A;Title: Cloning of a Clostridium thermoceum DNA fragment encoding polypeptides that

A;Reference number: S25767

A;Accession: S25767

A;Molecule type: DNA

A;Residues: 1216-1232,'Y',1234-1241,'T',1243-1319,'R',1321-1615,1617-1854 <FU2>

A;Cross-references: EMBL:X67406

R;Beguin, P.

submitted to the EMBL Data Library, August 1992

A;Reference number: S28659

A;Accession: S28659

A;Molecule type: DNA

A;Residues: 1216-1232,'Y',1234-1241,'T',1243-1319,'R',1321-1615,1617-1768,'R',1770-1854

R;Fujino, T.; Beguin, P.; Aubert, J.P.

J. Bacteriol. 175, 1891-1899, 1993

A;Title: Organization of a Clostridium thermoceum gene cluster encoding the celluloso

e.

A;Reference number: Z18847; MUID:93209931; PMID:8458832

A;Accession: T18261

A;Status: preliminary; translated from GB/EMBL/DBD

A;Molecule type: DNA

A;Residues: 1821-1854 <FU2>

A;Cross-references: EMBL:X67506; NID:g296879; PID:g296880; PIDN:CAA47840.1

C;Genetics:

A;Gene: cipA

Query Match 35.1%; Score 165; DB 2; Length 1854;

Best Local Similarity 36.1%; Pred. No. 6.7e-09;

Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVYNC 60

Db 365 VSNLKVVEFNPSPTTNSINPQKVTNTGSSAIDLKLTLYRYTYVDGQKDTFW-CD 423

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88

Db 424 HAAIIGSGSYNGITSNVKGTFVKMSSSTNNADTYLE 460

Search completed: October 5, 2005, 07:57:09

Job time : 3.46958 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:44:57 ; Search time 11.0656 Seconds
(without alignments)
4118.611 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGVKVQYKNDSPGDNQ.....RASFGSVNPATPTADTVLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	257	54.7	741	2 Q82QF2	Q82qf2 streptomyc
2	228.5	48.6	616	2 Q7X2N2	Q7x2n2 thermomono
3	225.5	48.0	170	2 Q9RFK6	Q9rfx6 caldibacill
4	216.5	46.1	930	2 Q9RFX5	Q9rfx5 caldibacill
5	214.5	45.6	921	2 Q9L8L8	Q9l8l8 caldibacill
6	209.5	44.6	1091	2 Q8KKF7	Q8kkf7 paenibacill
7	204.5	43.5	997	2 Q9Z4I1	Q9z4i1 bacillus sp
8	194.5	41.4	1751	2 Q9AQG4	Q9aqg4 caldicellul
9	192.5	41.0	1000	2 Q24820	O24820 thermophili
10	192.5	41.0	1770	2 Q9X3P5	Q9x3p5 caldicellul
11	191.5	40.7	261	2 Q9AQG7	Q9aqg7 caldicellul
12	191.5	40.7	1426	2 Q9X3P6	Q9x3p6 caldicellul
13	191	40.6	1711	2 P96311	P96311 anaerocellu
14	190.5	40.5	996	2 Q9AQH0	Q9aqh0 caldicellul
15	190.5	40.5	1779	2 Q52374	O52374 caldicellul
16	190	40.4	499	1 GUN1_BACSU	P07983 bacillus su
17	190	40.4	499	1 GUN2_BACSU	P10475 bacillus su
18	190	40.4	499	2 Q93TJ6	Q93tj6 bacillus su
19	190	40.4	508	2 Q93LD0	Q93ld0 bacillus su
20	190	40.4	508	2 Q70K05	Q70k05 bacillus am
21	188	40.0	499	1 GUN3_BACSU	P23549 bacillus su
22	187	39.8	499	2 Q52731	O52731 bacillus sp
23	187	39.8	499	2 Q674Q1	Q6y4q1 bacillus su
24	186.5	39.7	1331	1 MANB_CALSA	P22533 caldocellum
25	186.5	39.7	1742	1 GUNA_CALSA	P22534 caldocellum
26	186	39.6	499	2 Q45532	Q45532 bacillus su
27	186	39.6	501	2 Q83012	Q83012 bacillus sp
28	182.5	38.8	1039	1 GUNE_CALSA	P10474 c endogluc
29	180.5	38.4	145	1 YCEA_PAEIA	P29718 paenibacill
30	180	38.3	486	2 Q45430	Q45430 bacillus sp
31	177	37.7	499	2 Q8RPQ6	Q8rpq6 bacillus am

32 172.5 36.7 1915 2 Q9RPL0 acetivibrio
33 172 36.6 700 1 GUNA_PAEIA P29719 paenibacill
34 168 35.7 473 2 Q9RK75 streptomyc
35 165.5 35.2 879 1 GUN1_CLOTM Q02934 clostridium
36 165.5 35.2 887 2 Q9L3J8 clostridium
37 165 35.1 772 1 CIEB_CLOTM Q01866 clostridium
38 165 35.1 1853 1 CIEA_CLOTM Q06851 clostridium
39 157 33.4 534 2 Q65IY7 Q65iy7 bacillus li
40 157 33.4 542 2 Q7X3S6 Q7x3s6 bacillus li
41 150.5 32.0 2319 2 Q9FDJ9 Q9fdj9 bacteroides
42 149.5 31.8 444 1 GUNN_ERWCA O59394 erwinia car
43 149.5 31.8 505 1 GUNV_ERWCA Q47096 erwinia car
44 147.5 31.4 504 1 GUNW_ERWCA Q59395 erwinia car
45 147.5 31.4 914 1 GUX2_CLOSR P50900 clostridium

ALIGNMENTS

RESULT 1
Q82QF2 PRELIMINARY; PRT; 741 AA.
ID AC Q82QF2
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Putative cellulose 1,4-beta-cellobiosidase.
GN Name=guxal; OrderedLocustNames=SAV557;
OS Streptomycetes avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonce T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,
RT "Genome sequence of an industrial microorganism Streptomycetes
RT avermitilis; deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.,
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomycetes avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
DR EMBL; AP005023; BAC68267.1; -.
DR HSSP; P20533; 1K85
DR GO; GO:0004553; Phydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; Polysaccharide metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR003961; FN_III-like.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR01524; Glyco_hydro_6.
DR InterPro; IPR011253; Glyco_hydro_6-1k.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLHYDRLASE6.
DR ProDom; PD001947; CBD_3; 1.
DR ProDom; PD003733; Glyco_hydro_6; 1.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS00853; FN3; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
KW Complete proteome.

QY	2	SGGVKQVYKNDSSAPGDNQIKPGQLQVNTGSSVDLSVTIVRYWFTRDGGSSSTLVYNCDW	61
DB	591	SGGLKVLYKNDSSATDINAIRPGLRIVNTGSLDLISKVTARYFYTRDSGSTVNACDY	650
QY	62	AAMCGCNIRASFGSVNPATPTADTYLQ	88
DB	651	AAVGCSNVSKVLPTTPVECADAYLE	677
 RESULT 2 Q7X2N2 PRELIMINARY; PRT; 616 AA.			
ID	Q7X2N2	PRELIMINARY;	PRT; 616 AA.
AC	Q7X2N2;		
DT	01-OCT-2003 (TtEMBLrel. 25, Created)		
DT	01-OCT-2003 (TtEMBLrel. 25, Last sequence update)		
DT	01-MAR-2004 (TtEMBLrel. 26, Last annotation update)		
DE	Endoglucanase.		
GN	Name=cel5B;		
OS	Thermomonospora fusca.		
OC	Bacteria; Actinobacteria; Actinomycetales;		
OC	Streptosporangineae; Nocardiopsaceae; Thermobifida.		
NCBI_TaxID=2021;	[1]_RN		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=TW51;		
RC	Posta K., Beki E., Kukolya J., Hornok L.;		
RL	Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY298814; AAP56348.1; -.		
DR	HSP; Q45996; IG43		
GO; GO:0004553; P:	hydrolyzing O-glycosyl . . . ; IEA.		
GO; GO:0005975; P:	carbohydrate metabolism; IEA.		
InterPro; IPR001956; CDB 3.			
InterPro; IPR008965; Cellul bind.			
InterPro; IPR001547; Glyco_Hydro_5.			
Pfam; PF00942; CBM_3; 1..			
Pfam; PF00150; Cellulase; 1.			
ProDom; PD001947; CBD_3; 1.			
PROSITE; PS00659; GLYCOSYL HYDROL_F5; UNKNOWN 1.			
SEQUENCE	616 AA; 67701 MW; 24FFC1EA1A3F5639 CRC64;		
 Query Match 48.6%; Score 228.5; DB 2; Length 616; Best Local Similarity 52.3%; Pred. No. 1.7e-16; Matches 45; Conservative 16; Mismatches 22; Indels 3; Gaps 2;			
QY	2	SGGVKQVYKNDSSAPGDNQIKPGQLQVNTGSSVDLSVTIVRYWFTRDGGSSSTLVYNCDW	61
DB	471	TGALEVYVRNNLSAADSQIAPGLRVNTGSSVTDLADVIEHYFTNEPG-GTLQFTCDW	529
QY	62	AAMCGCNIRASFGSVNPATPTADTYL	87
DB	530	AQVGCAVNASFTSL--SAPGADTSL	553
 RESULT 3 Q9RFX6 PRELIMINARY; PRT; 170 AA.			
AC	Q9RFX6;		
DT	01-MAY-2000 (TtEMBLrel. 13, Created)		
DT	01-MAY-2000 (TtEMBLrel. 13, Last sequence update)		
DT	01-OCT-2003 (TtEMBLrel. 25, Last annotation update)		
DE	Hypothetical protein (Fragment).		
OS	Caldibacillus cellulovorans.		
OC	Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldibacillus.		
NCBI_TaxID=74586;	[1]		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=20120520; PubMed=10653733; DOI=10.1128/AEM.66.2.664-670.2000;		

RA	Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT	"A gene encoding a novel multidomain beta-1,4-mannanase from
RT	Caldibacillus cellulovorans and action of the recombinant enzyme on
RT	kraft pulp.";
RL	Appl. Environ. Microbiol. 66:664-670(2000).
DR	EMBL; AF163837; AAF22273.1; --
DR	HSSP; QO6851; INEC.
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR	InterPro; IPR001956; CBD_3.
DR	InterPro; IPR008965; Cellul_bind.
DR	Pfam; PF00942; CBM_3; 1.
DR	ProDom; PD001947; CBD_3; 1.
DR	Hypothetical protein.
KW	NON_TER
FT	1
SQ	SEQUENCE 170 AA; 18493 MW; 7AC9D33F44E3A0B4 CRC64;
	Query Match 48.0%; Score 225.5; DB 2; Length 170;
	Best Local Similarity 47.7%; Pred. No. 8.3e-17;
	Matches 41; Conservative 17; Mismatches 27; Indels 1; Gaps
QY	3 GGKVVQYKNDSAPGDNQIKPGQLQVNTGSSSVDLSTVTTRYWFTRDGSGSLTVINCDA 62
	: :
Db	20 GSLVQVYRAADTNAGDNQKLPHFRIVNRGTSSVPISLTI RYYITVD-GDKPQVENCDWA 78
QY	63 AMCGGNIRASFGSVNPATPTADTYLQ 88
	: : : : : : : : : : : : : : : :
Db	79 QVGCSNVRGSFVKLSLTGRTGADYIIE 104
RESULT 4	
Q9RFX5	PRELIMINARY; PRT; 930 AA.
ID	Q9RFX5 PRELIMINARY; PRT; 930 AA.
AC	O9REX5;
DT	01-MAY-2000 (TrEMBLrel. 13, Created)
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE	Multidomain beta-1,4-mannanase precursor.
GN	Name=mana;
OS	Caldibacillus cellulovorans.
OC	Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldibacillus.
OX	NCBI_TaxID=74596;
RN	[1]
RP	SSEQUENCE FROM N.A.
EX	MEDLINE=20120520; PubMed=10653733; DOI=10.1128/AEM.66.2.664-670.2000;
RA	Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT	"A Gene encoding a novel multidomain beta-1,4-mannanase from
RT	Caldibacillus cellulovorans and action of the recombinant enzyme on
RT	kraft pulp.";
RL	Appl. Environ. Microbiol. 66:664-670(2000).
DR	EMBL; AF163837; AAF22274.1; --
DR	HSSP; Q9ZF13; 3MAN.
DR	GO; GO:0019028; C:viral capsid; IEA.
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . ; IEA.
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.

Query Match 46.1%; Score 216.5; DB 2; Length 930;
Best Local Similarity 46.5%; Pred. No. 5.9e-15;
Matches 40; Conservative 17; Mismatches 28; Indels 1; Gaps 1;
OV 3 GGKVKYQKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVYWFTRDGGSSTLVYNCDA 62

Db 780 GNLVQVRAADTNATDNQLKPHFRIVNRGTSVPLSBLTIRYWTVD-GDKPQVFNCDWA 838
Qy 63 AMGCNIRASFGSVNPAFTADTYLQ 88
Db 839 QVGCNLRGSFVKLSLSTGRTGADYYIE 864

RESULT 5
ID Q9L8L8 PRELIMINARY; PRT; 921 AA.
AC Q9L8L8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Beta-1,4-xylanase XynA precursor.
GN Name=xynA;
OS Caldicellulosum cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicellulosum.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20519260; PubMed=11065373;
RA Sunna A., Gibbs M.D., Bergquist P.L.;
RT "A novel thermostable multigene family 1,4-beta-xylanase from
RT 'Caldicellulosum cellulovorans' and effect of its xylan-binding domain
RT on enzyme activity."
RL Microbiology 146:2947-2955(2000).
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AF200304; AAP61649.1; -.
DR HSSP; Q06851; INCB.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR00979; Gal_bind_like.
DR InterPro; IPR01000; Glyco_hydro_10.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF02018; CBM_4; 9; 1.
DR PRINTS; PF00331; Glyco_hydro_10; 1.
DR PRODOM; PD001947; CBD_3; 2.
DR SMART; SM00633; Glyco_10; 1.
DR GLYCOSIDASE; Hydrolase; Signal; Xylan degradation.
FT SIGNAL 1 33 Potential.
FT CHAIN 34 921 beta-1,4-xylanase XynA.
SQ SEQUENCE 921 AA; 102380 MW; C5DDDIATF7567413 CRC64;

Query Match 45.6%; Score 214.5; DB 2; Length 921;
Best Local Similarity 46.5%; Pred. No. 9,7e-15;
Matches 40; Conservative 16; Mismatches 29; Indels 1; Gaps 1;

Qy 3 GGVKQVKNDSAPGDNQIKFGLQVNTGSSVSLSTVTVYWFTRDGSSTLVNCDWA 62
Db 771 GNLVQVRAADTNAGDNQLKPHFRIVNRGTSVPLSLTIRYWTVD-GDKPQVFNCDWA 829

Qy 63 AMGCNIRASFGSVNPAFTADTYLQ 88
Db 830 WVGCSNLRGSLVKLTGRTGADYYLE 855

RESULT 6
ID Q8KKF7 PRELIMINARY; PRT; 1091 AA.
AC Q8KKF7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulose 1,4-beta-cellobiosidase precursor (EC 3.2.1.91).
GN Name=cel48C;
OS Paenibacillus sp. BP-23.

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=198119;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-23;
RA Sanchez M.M., Pastor F.I.J., Diaz P.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 2 fibronectin type III domains.
DR EMBL; AJ488933; CAD32945.1; -.
DR HSSP; P36866; 1LIV.
DR GO; GO:0016162; F:cellulose 1,4-beta-cellobiosidase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00041; fn3; 2.
DR PRINTS; PF02011; Glyco_hydro_48; 1.
DR PRODOM; PD001947; CBD_3; 1.
DR PRODOM; PD011903; Glyco_hydro_48; 2.
DR SMART; SM0060; FN3; 2.
DR PROSITE; PS00853; FN3; 2.
DR GLYCOSIDASE; Hydrolase; Signal.
FT SIGNAL 1 35 Potential.
FT CHAIN 36 1091 cellulose 1,4-beta-cellobiosidase.
SQ SEQUENCE 1091 AA; 118000 MW; 21EACCEB2E704478 CRC64;

Query Match 44.6%; Score 209.5; DB 2; Length 1091;
Best Local Similarity 46.0%; Pred. No. 4.2e-14;
Matches 40; Conservative 21; Mismatches 25; Indels 1; Gaps 1;

Qy 2 SGGVKQVKNDSAPGDNQIKFGLQVNTGSSVSLSTVTVYWFTRDGSSTLVNCDW 61
Db 940 TGTLEVOYENGSGSASGNATPQFNLKNTGTTAIDLSKVRYFTKD-SADMSFWCDY 998

Qy 62 AAMGCNIRASFGSVNPAFTADTYLQ 88
Db 999 AQLGSANVQGSFVAVNPAPKGTADTYLE 1025

RESULT 7
ID Q9Z4I1 PRELIMINARY; PRT; 997 AA.
AC Q9Z4I1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cellulase precursor (EC 3.2.1.4).
GN Name=celB;
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RL Molecular cloning and characterization of a multidomain endoglucanase
RL from Paenibacillus sp BP-23: evaluation of its performance in pulp
RL refining."
RL Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -!- SIMILARITY: Contains 2 fibronectin type III domains.
DR EMBL; AJ13614; CAB38941.1; -.
DR HSSP; P26221; 1TF4.
DR GO; GO:0008810; F:cellulase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR003961; FN_III.


```

DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR002345; Lipocalin.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 2.
DR SMART; SM00060; FN3_2.
DR PROSITE; PS00853; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34 Potential.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 43.5%; Score 204.5; DB 2; Length 997;
Best Local Similarity 43.7%; Pred. No. 1.3e-13;
Matches 38; Conservative 21; Mismatches 27; Indels 1; Gaps 1;

QY 2 SGVKVQYKNDSPAGDNOIKPGLQVNTGSSVDLSITVTVRYWFTRDGGSSTLVYNCW 61
DB 846 TGTEVQYRSGSGSSNAVTPQFNKNTGTQADLSTVKIRYFTKD-GTEELSFWDY 904

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
DB 905 AQGSANVQGMFVAVNPAKGTADTYVE 931

RESULT 8
Q9AQA4 PRELIMINARY; PRT; 1751 AA.
AC Q9AQA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078042; AAK06394.1; -.
DR HSSP; P37700; 1G87.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR002345; Lipocalin.
DR Pfam; PF00942; CBM_3; 4.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 4.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 41.4%; Score 194.5; DB 2; Length 1751;
Best Local Similarity 44.7%; Pred. No. 3.3e-12;
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNOIKPGLQVNTGSSVDLSITVTVRYWFTRDGGSSTLVYNCWAA 63
DB 846 TGTEVQYRSGSGSSNAVTPQFNKNTGTQADLSTVKIRYFTKD-GTEELSFWDY 904

Db 678 GVKVLYKNNETSASTGSRPMFKIVNGSSSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 736
QY 64 MCGNIRASFGSVNPATPTADTYLQ 88
DB 737 IGASNVTFNFVKLSGSGVSGADYYLE 761

RESULT 9
Q24820 PRELIMINARY; PRT; 1000 AA.
AC Q24820;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RL Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AB008029; BAA22939.1; -.
DR HSSP; Q06851; 1NEC.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; Cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1000 AA; 113264 MW; B9F659A56A752C6B CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity 43.7%; Pred. No. 2.9e-12;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGVKVQYKNDSPAGDNOIKPGLQVNTGSSVDLSITVTVRYWFTRDGGSSTLVYNCW 61
DB 372 SQGKIVLYANKETNSTTIRPWLKVNTGSSSIDLSRVITRYWFTVDGDKAQSAYS-DW 430

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
DB 431 AQIGASNVTFNFVKLSGSGVSGADYYLE 457

RESULT 10
Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC Q9X3P5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Xyna.
GN Name-xyna;
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]

```



```
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AF078737; AAD30363.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR006584; CBD_IV.
DR InterPro; IPR005084; CBM_6.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR006710; Glyco_hydro_43.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02018; CBM_4; 9; 2.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR Pfam; PF04616; Glyco_hydro_43; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00606; CBD_IV; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1770 AA; 193640 MW; 8BAF1937D4926C92 CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1770;
Best Local Similarity 43.5%; Pred. No. 5.6e-12;
Matches 37; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

QY 4 GVKVQYKNDGAPGDNQIKPGLQLVNTGSSVDLSTVTYVYWFTRDGGSSSTLVYNCDAWAA 63
Db 1104 GVKVYKNNETSASTSSIRPFKIVNGSSVDLSRVKIRYVTVGDKPQSAV-CDWAAQ 1162
QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88
Db 1163 IGASVNTFNFVKLSGSGVGDYYLE 1187

RESULT 11
Q9AQG7 PRELIMINARY; PRT; 261 AA.
AC Q9AQG7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 6 (Fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078040; AAK06391.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
```

```
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
KW Hydrolase.
FT NON_TER 1 261
FT NON_TER 261
SQ SEQUENCE 261 AA; 28759 MW; 4771744A26A6AE04 CRC64;

Query Match 40.7%; Score 191.5; DB 2; Length 261;
Best Local Similarity 43.5%; Pred. No. 7.7e-13;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDGAPGDNQIKPGLQLVNTGSSVDLSTVTYVYWFTRDGGSSSTLVYNCDAWAA 63
Db 118 GLKVLVYKNNETSASTSSIRPFKIVNGSSVDLSRVKIRYVTVGDKPQSAV-CDWAAQ 176
QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88
Db 177 IGASVNTFNFVKLSGSGVGDYYLE 201

RESULT 12
Q9X3P6 PRELIMINARY; PRT; 1426 AA.
ID Q9X3P6;
AC Q9X3P6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE CellB.
DE Name=cellB;
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AF078737; AAD30364.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF00150; Cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1426 AA; 157543 MW; 29B3FDB85D09A863 CRC64;

Query Match 40.7%; Score 191.5; DB 2; Length 1426;
Best Local Similarity 43.5%; Pred. No. 5.6e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDGAPGDNQIKPGLQLVNTGSSVDLSTVTYVYWFTRDGGSSSTLVYNCDAWAA 63
```



```

Db 413 GLKLVYKNNETSASTGSRPFVKIYNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 471
QY 64 MCGNIRASFGSVNPAFTADTYLQ 88
Db 472 IGASNVTFNFVKLSSVSGADYILE 496

RESULT 13
P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE 1,4-beta-glucanase (Fragment).
GN Name=CelA;
OS Anaerocellum thermophilum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Anaerocellum group;
OC Anaerocellum.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z-1320;
RX MEDLINE=98154434; PubMed=9493383;
RA Zverlov V., Mahr S., Riedel K., Bronnenmeier K.;
RT "Properties and gene structure of a bifunctional cellulolytic enzyme
RT (CelA) from the extreme thermophile Anaerocellum thermophilum with
RT separate glycosyl hydrolase family 9 and 48 catalytic domains.";
RL Microbiology 144:457-465(1998).
DR EMBL; Z86105; CAB06786.1; -.
DR PIR; T31337; T31337.
DR HSSP; P37700; 1687.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR000556; Glyco_hydro_48.
DR InterPro; IPR003701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR000209; Pept_S8_S53.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom; PD001947; CBD_3; 3.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1711 AA; 169979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 40.6%; Score 191; DB 2; Length 1711;
Best Local Similarity 43.8%; Pred. NO. 7.8e-12;
Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

QY 1 VSGG-VKVOYKNDSPAGDNOIKPGLQVNTGSSVDLSVTYVYWFTRDGGSSTLVYNC 59
Db 683 VAGQIKVLYANKNTSTNTIRPWLKVNTGSSSIDLSRVTIIRYWTVDGDKAQSALS- 741

QY 60 DWAAAMCGNIRASFGSVNPAFTADTYLQ 88
Db 742 DWAGIGASNVTFKFKVLLSSVSGADYILE 770

RESULT 14
Q9AQHO PRELIMINARY; PRT; 996 AA.
AC Q9AQHO;
AC Q9AQHO;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

```

```

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (Fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSSP; Q06851; INEC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR Hydrolase. 996 996
KW NON_TER
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match 40.5%; Score 190.5; DB 2; Length 996;
Best Local Similarity 43.5%; Pred. No. 4.7e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVOYKNDSPAGDNOIKPGLQVNTGSSVDLSVTYVYWFTRDGGSSTLVYNC 63
Db 844 GLKLVYKNNETSASAGSRPFVKIYNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 902

QY 64 MCGNIRASFGSVNPAFTADTYLQ 88
Db 903 IGASNVTFNFVKLSSVSGADYILE 927

RESULT 15
OS2374 PRELIMINARY; PRT; 1779 AA.
ID OS2374;
AC OS2374;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN Name=xynC;
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rt69B.1;
RC MEDLINE=99283888; PubMed=10356996;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RA "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. strain
RT Rt69B.1.";
RL Extremophiles 3:103-111(1999).
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
CC EMBL; AF036924; AAB95326.1; -.
DR PIR; T31085; T31085.
DR HSSP; Q06851; INBC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR006584; CBD_IV.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:49:28 ; Search time 3.41942 Seconds
(without alignments)
1942.949 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQVQKNDSPGDNQ.....RASFGVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/5A-COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195.5	41.6	616	3	US-09-136-574A-47
2	194.5	41.4	1751	3	US-09-136-574A-44
3	191.5	40.7	1428	3	US-09-136-574A-43
4	172	36.6	700	2	US-07-862-588B-2
5	165.5	35.2	551	2	US-09-033-537A-1
6	165	35.1	167	5	PCT-US95-13813-9
7	165	35.1	476	4	US-09-339-159B-4
8	165	35.1	493	3	US-09-198-956-10
9	165	35.1	493	3	US-09-198-955A-12
10	165	35.1	493	3	US-09-694-531-12
11	165	35.1	493	4	US-09-670-141-10
12	165	35.1	493	4	US-10-072-153-12
13	155.5	33.1	1352	4	US-09-784-554B-2
14	150.5	32.0	1350	4	US-09-784-554B-4
15	120.5	25.6	531	2	US-07-862-588B-7
16	114	24.3	162	1	US-08-048-164A-2
17	114	24.3	162	1	US-08-460-463-2
18	114	24.3	162	1	US-08-460-457-2
19	114	24.3	162	1	US-08-460-458-2
20	114	24.3	162	2	US-08-460-455-2
21	114	24.3	162	2	US-08-330-394A-2
22	114	24.3	163	3	US-09-006-636-7
23	114	24.3	163	3	US-09-006-632-7
24	114	24.3	163	3	US-09-325-274-7
25	113	24.0	382	3	US-09-277-716-22
26	113	24.0	382	4	US-09-609-161B-22
27	112	23.8	154	2	US-08-330-394A-29

28	112	23.8	156	2	US-08-330-394A-22	Sequence 22, Appl
29	72.5	15.4	272	4	US-09-902-540-11056	Sequence 11056, A
30	65.5	13.9	1566	4	US-09-581-472B-2	Sequence 2, Appl1
31	65	13.8	474	4	US-09-248-796A-25524	Sequence 25524, A
32	64	13.6	143	4	US-09-301-593-26	Sequence 26, Appl
33	64	13.6	428	3	US-09-118-319-5	Sequence 5, Appli
34	64	13.6	464	1	US-08-353-400-36	Sequence 36, Appl
35	64	13.6	472	4	US-09-301-593-30	Sequence 30, Appl
36	63.5	13.5	1785	3	US-09-341-587-3	Sequence 18, Appl
37	63	13.4	453	4	US-09-301-593-18	Sequence 43, Appl
38	63	13.4	472	4	US-09-301-593-43	Sequence 4640, Ap
39	63	13.4	718	4	US-09-328-352-4640	Sequence 12243, A
40	63	13.4	1133	4	US-09-902-540-12243	Sequence 2, Appl1
41	63	13.4	1581	3	US-09-110-517-2	Sequence 38, Appl
42	62.5	13.3	288	3	US-09-423-439-38	Sequence 14347, A
43	62.5	13.3	389	4	US-09-902-540-14347	Sequence 33, Appl
44	62.5	13.3	445	1	US-08-353-400-33	Sequence 8, Appli
45	62.5	13.3	641	4	US-09-687-538B-8	

ALIGNMENTS

RESULT 1
US-09-136-574A-47
; Sequence 47, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 616 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-09-136-574A-47

Query Match 41.6%; Score 195.5; DB 3; Length 616;
 Best Local Similarity 43.2%; Pred. No. 3.5e-13; Indels 1; Gaps 1;
 Matches 38; Conservative 15; Mismatches 34;
 QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGKPKQSAV-CD 59
 Db 1 MGSQVLYKNETSASTGSRPFKIVNGSSVDLSRVKIRYWTVDGKPKQSAV-CD 59
 QY 61 WAAMCGNIRASFGSVNPTATDTYLQ 88
 Db 60 WAQIGASNVTFNFKLSSGSGADYYLE 87

RESULT 2

US-09-136-574A-44
 ; Sequence 44, Application US/09136574A
 ; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.
 Anderson, Paige
 Gibbs, Moreland
 Bergquist, Peter
 Daniels, Roy
 Morgan, Hugh W.
 Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
 Treating Cellulose Containing Fabrics Using Truncated
 Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
 STREET: Spring House Corporate Center, P.O. Box 457
 CITY: Spring House
 STATE: PA
 COUNTRY: USA
 ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A
 FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571
 FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.
 REGISTRATION NUMBER: 31,215
 REFERENCE/DOCKET NUMBER: 1997US001/CIP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-540-9200
 TELEFAX: 215-540-5818
 TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1751 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44
 Query Match 41.4%; Score 194.5; DB 3; Length 1751;
 Best Local Similarity 44.7%; Pred. No. 1.6e-12;
 Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGKPKQSAV-CD 63

Db 678 GVKVLYKNETSASTGSRPFKIVNGSSVDLSRVKIRYWTVDGKPKQSAV-CD 736

QY 64 MCGNIRASFGSVNPTATDTYLQ 88
 Db 737 IGASNVTFNFKLSSGSGADYYLE 761

RESULT 3

US-09-136-574A-43

; Sequence 43, Application US/09136574A
 ; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.
 Anderson, Paige
 Gibbs, Moreland
 Bergquist, Peter
 Daniels, Roy
 Morgan, Hugh W.
 Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
 Treating Cellulose Containing Fabrics Using Truncated
 Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
 STREET: Spring House Corporate Center, P.O. Box 457
 CITY: Spring House
 STATE: PA
 COUNTRY: USA
 ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A
 FILING DATE: 19-Aug-1998
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571
 FILING DATE: September 19, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Bak, Mary E.
 REGISTRATION NUMBER: 31,215
 REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200
 TELEFAX: 215-540-5818
 TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1426 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: No. 6294366e

SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43
 Query Match 40.7%; Score 191.5; DB 3; Length 1426;
 Best Local Similarity 43.5%; Pred. No. 2.7e-12;
 Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGKPKQSAV-CD 63

Db 413 GLKVLKNETSASTGSRPFKIVNGSSVDLSRVKIRYWTVDGKPKQSAV-CD 471

QY 64 MCGNIRASFGSVNPTATDTYLQ 88

Db 472 IGASNVTFNFKLSSGSGADYYLE 496

RESULT 4

STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10174

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/033.537A
FILING DATE: 02-MAR-1998
CLASSIFICATION: 008
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 0993/95
FILING DATE: 08-SEP-1995
APPLICATION NUMBER: PCT/DK96/00364
FILING DATE: 03-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4492.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-033-537A-1

Query Match 35.2%; Score 165.5; DB 2; Length 551;
Best Local Similarity 37.9%; Pred. No. 6e-10;
Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;

DDB 2 SGGVQVKYKNNDSPAGDNIQPLQLVNNTGSSSDLSLTVRYWFTRDGSSTLVYNCW 61
 :
 402 TGNLVQVKVGDTSATDNQMKEFNFKNGTTPVNLSQLKLRYFFTKD-GTADWSASFWD 460

DDB 62 AAMGCNIRASFGSVNPATPTADTYLQ 88
 :
 461 AQIGASNVSAAF--ANFTGSNTDTYE 485

RESULT 6
PCT-US95-13813-9
Sequence 9, Application PC/TUS9513813
GENERAL INFORMATION:
APPLICANT: Yeda Research and Development Co. Ltd.
APPLICANT: Ramot University Authority for Applied
APPLICANT: Research and Industrial Development Ltd.
APPLICANT: Technion Research and Development Foundation Ltd.
APPLICANT: Bayer, Edward A.
APPLICANT: Morag, Ely
APPLICANT: Wilchek, Meir
APPLICANT: Lamed, Raphael
APPLICANT: Shoham, Yuval
TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
TITLE OF INVENTION: PROTEINS AND USE THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSES: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

US-07-862-588B-2
Sequence 2, Application US/07862588B
Patent No. 5916796

GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linas
APPLICANT: Schlein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIORITY APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein

US-07-862-588B-2

Query Match 36.6%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred.No.1.6e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3

QY 1 VSGGVKYQQKNND-SAPGDNQIKPGQLVNTGSSVDLSIVTVRYWFTEDGGSSLVLVNC 59
DDB : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
 549 VNDSLVRYYQGDKGRNATDNQIKPHFNQNKGTSVPDLISLTLYRYYFKTD--SAAAMNGWI 607

QY 60 DWAAAGCGCNIRASFGSVNPATPTADTYLQ 88
DDB 608 DWAUKGSNIQSIFGNHGA--DSDTYAE 634

RESULT 5
US-09-033-537A-1
Sequence 1, Application US/09033537A
Patent No. 5958083

GENERAL INFORMATION:
APPLICANT: Onishi, Masahiro
APPLICANT: Fich, Merete
APPLICANT: Toft, Annette Hanne
APPLICANT: Shlein, Martin
TITLE OF INVENTION: Prevention Of Back-Staining
TITLE OF INVENTION: In Stone Washing
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. 5958083O NO. 5958083disk of NO. 5958083th America, Inc.


```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: BAYER-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 167 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-13813-9

Query Match 35.1%; Score 165; DB 5; Length 167;
Best Local Similarity 36.1%; Pred. No. 1.6e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVQKNDSSAPGDNQIKPGLQLVNTGSSVDLSTVTVRVYWFTRDGGSSSTLVNCD 60
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 5 VSGNLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTIRLYYYTVDGQKDTFW-CD 63

QY 61 WAAM-----GCGNIRASFGSVNPTPTADTYLQ 88
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 64 HAAIIGNSYNGITSNVKGTFTVKMSSSTNNADTYLE 100

RESULT 7
US-09-339-159B-4
; Sequence 4, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Bacillus
US-09-339-159B-4

Query Match 35.1%; Score 165; DB 4; Length 476;
Best Local Similarity 36.1%; Pred. No. 5.7e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVQKNDSSAPGDNQIKPGLQLVNTGSSVDLSTVTVRVYWFTRDGGSSSTLVNCD 60
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 314 VSGNLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTIRLYYYTVDGQKDTFW-CD 372

QY 61 WAAM-----GCGNIRASFGSVNPTPTADTYLQ 88
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 373 HAAIIGNSYNGITSNVKGTFTVKMSSSTNNADTYLE 409

RESULT 8
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-10

Query Match 35.1%; Score 165; DB 3; Length 493;
Best Local Similarity 36.1%; Pred. No. 6e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVQKNDSSAPGDNQIKPGLQLVNTGSSVDLSTVTVRVYWFTRDGGSSSTLVNCD 60
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 331 VSGNLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTIRLYYYTVDGQKDTFW-CD 389

QY 61 WAAM-----GCGNIRASFGSVNPTPTADTYLQ 88
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 390 HAAIIGNSYNGITSNVKGTFTVKMSSSTNNADTYLE 426

RESULT 9
US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
```



```
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match      35.1%; Score 165; DB 3; Length 493;
Best Local Similarity 36.1%; Pred. No. 6e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 60
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 331 VSGNLKVEFYNSPDDTTNSINPQFKVTYTGSSAIDLKLTLYYYTVDGQKQTFW-CD 389
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQ 88
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 390 HAAIGSGSYNGITSNVKGTFFVKMSSTNNADTYLE 426
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 10
US-09-694-531-12
; Sequence 12, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-694-531-12

Query Match      35.1%; Score 165; DB 3; Length 493;
Best Local Similarity 36.1%; Pred. No. 6e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 60
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 331 VSGNLKVEFYNSPDDTTNSINPQFKVTYTGSSAIDLKLTLYYYTVDGQKQTFW-CD 389
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQ 88
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 390 HAAIGSGSYNGITSNVKGTFFVKMSSTNNADTYLE 426
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 11
US-09-670-141-10
; Sequence 10, Application US/09670141
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-670-141-10

Query Match      35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 6e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSTLVNCD 60
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 331 VSGNLKVEFYNSPDDTTNSINPQFKVTYTGSSAIDLKLTLYYYTVDGQKQTFW-CD 389
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQ 88
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 390 HAAIGSGSYNGITSNVKGTFFVKMSSTNNADTYLE 426
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 12
US-10-072-152-12
; Sequence 12, Application US/10072152
; Patent No. 6677147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6677147el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; PRIOR FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
```



```

; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-10-072-152-12

Query Match      35.1%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 6e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60
Db 331 VSGNLKVEFYNSPDDTNSINPOKVTNTGSSAIDLSKLTIRYYTVDGQKQTFW-CD 389

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIIGNSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

RESULT 13
US-09-784-554B-2
; Sequence 2, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1352
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-2

Query Match      33.1%; Score 155.5; DB 4; Length 1352;
Best Local Similarity 31.8%; Pred. No. 2.3e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAA 63
Db 1200 GLLQYRTADTKVNDNHLNPHQIIVNKGTSPINELKIRYYTIDGDREQ-TFNC DYAT 1258

QY 64 MCGNIRASFGSVNPATPTADTYLQ 88
Db 1259 LSCSKLNGKLVKMKAKATGADYYLE 1283

RESULT 14
US-09-784-554B-4
; Sequence 4, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1350
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-4

Query Match      32.0%; Score 150.5; DB 4; Length 1350;

```

```

Best Local Similarity 31.8%; Pred. No. 8.1e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAA 63
Db 1199 GLVQYRTADTKVNDNHLNPHQIIVNKGTSPINELKIRYYTIDGDREQ-TFNC DYAV 1257

QY 64 MCGNIRASFGSVNPATPTADTYLQ 88
Db 1258 LSCSKLNGKLVKMKAKATGADYYLE 1282

RESULT 15
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796ch America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-7

Query Match      25.6%; Score 120.5; DB 2; Length 531;
Best Local Similarity 32.9%; Pred. No. 5e-05;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

QY 2 SGGVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDW 61
Db 432 TGNLVVQYKVGDTSATDNQKPSFNKNGTTPVNLGKLVKLVKLVKLVKLVKLVKLV 490

QY 62 AAMGCGNIRASFGSVNPATPTADTY 86
Db 491 AQIGRTNVLLAF--ANFTGSGNTDITY 513

Search completed: October 5, 2005, 08:04:31

```


Job time : 4.41942 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:52:23 ; Search time 11.3031 Seconds
(without alignments)
3268.602 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQVYKNNDSAPGDNQ.....RASFGVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues
Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	469	99.8	88	11	US-09-917-376-5
2	469	99.8	88	14	US-10-155-400-5
3	469	99.8	89	11	US-09-917-376-4
4	469	99.8	89	14	US-10-155-400-4
5	469	99.8	154	10	US-09-917-378-4
6	469	99.8	762	10	US-09-917-378-1
7	466	99.1	150	10	US-09-917-384-5
8	466	99.1	150	10	US-09-917-383-5
9	466	99.1	1043	10	US-09-917-384-6
10	466	99.1	1043	10	US-09-917-383-6
11	466	99.1	1228	10	US-09-917-384-1

12	466	99.1	1228	10	US-09-917-383-1	Sequence 1, Appli
13	462	98.3	957	11	US-09-917-376-1	Sequence 1, Appli
14	462	98.3	957	14	US-10-155-400-1	Sequence 1, Appli
15	257	54.7	741	14	US-10-156-761-8100	Sequence 8100, Ap
16	190	40.4	508	15	US-10-369-493-23151	Sequence 23151, A
17	188	40.0	477	16	US-10-466-208-12	Sequence 12, Appli
18	188	40.0	496	16	US-10-466-208-8	Sequence 8, Appli
19	188	40.0	677	16	US-10-433-577-35	Sequence 35, Appli
20	180	38.3	1621	14	US-10-185-990-10	Sequence 10, Appli
21	165	35.1	167	17	US-10-933-404-4	Sequence 4, Appli
22	165	35.1	476	15	US-10-372-054-4	Sequence 4, Appli
23	165	35.1	493	13	US-10-072-152-12	Sequence 12, Appli
24	165	35.1	493	15	US-10-655-433-12	Sequence 12, Appli
25	165	35.1	599	10	US-09-955-555A-29	Sequence 29, Appli
26	155.5	33.1	1352	10	US-09-784-554B-2	Sequence 2, Appli
27	155.5	33.1	1352	16	US-10-896-555-2	Sequence 2, Appli
28	150.5	32.0	1350	10	US-09-784-554B-4	Sequence 4, Appli
29	150.5	32.0	1350	16	US-10-896-555-4	Sequence 4, Appli
30	130.5	27.8	1483	15	US-10-282-122A-51483	Sequence 51483, A
31	114	24.3	256	14	US-10-261-446-6	Sequence 6, Appli
32	114	24.3	256	15	US-10-261-445B-6	Sequence 6, Appli
33	114	24.0	256	17	US-10-782-234-6	Sequence 6, Appli
34	113	24.0	382	10	US-09-808-898-22	Sequence 22, Appli
35	105.5	22.4	163	15	US-10-460-524-2	Sequence 2, Appli
36	78.5	16.7	1049	15	US-10-282-122A-49900	Sequence 49900, A
37	75.5	16.1	346	16	US-10-437-963-179234	Sequence 223, App
38	75.5	16.1	618	15	US-10-211-462-223	Sequence 223, App
39	75.5	16.1	618	16	US-10-723-860-4039	Sequence 4039, Ap
40	75.5	16.1	618	18	US-10-756-149-5713	Sequence 5713, Ap
41	75.5	16.1	986	9	US-09-747-835A-53	Sequence 53, Appli
42	75.5	16.1	986	14	US-10-120-604-101	Sequence 101, App
43	75.5	16.1	986	14	US-10-225-567A-406	Sequence 406, App
44	75.5	16.1	986	15	US-10-398-458-3	Sequence 3, Appli
45	75.5	16.1	986	15	US-10-312-312-53	Sequence 53, Appli

ALIGNMENTS

RESULT 1
US-09-917-376-5
; Sequence 5, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 5
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-5

Query Match	99.8%	Score 469;	DB 11;	Length 88;
Best Local Similarity	100.0%;	Pred. No. 3.6e-48;		
Matches	88;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
OY	1	VSGGVKQVYKNNDSAPGDNQIKFGLQLVNTGSSVDLSTVTYVYWFTRDGGSSFLVYVNC	60	
DB	1	VSGGVKQVYKNNDSAPGDNQIKFGLQLVNTGSSVDLSTVTYVYWFTRDGGSSFLVYVNC	60	
OY	61	WRAMGCGNIRASFGSVNPATPTADTYLQ	88	

Db	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
RESULT 2			
US-10-155-400-5			
; Sequence 5, Application US/10155400			
; Publication No. US20030108988A1			
; GENERAL INFORMATION:			
; APPLICANT: DING, SHI-YOU			
; APPLICANT: ADNEY, WILLIAM S.			
; APPLICANT: VINZANT, TODD B.			
; APPLICANT: HIMMEL, MICHAEL E.			
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS			
; TITLE OF INVENTION: CELLULOXYTICUS			
; FILE REFERENCE: NREL 01-36A			
; CURRENT APPLICATION NUMBER: US/10/155,400			
; CURRENT FILING DATE: 2002-10-22			
; NUMBER OF SEQ ID NOS: 7			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 5			
; LENGTH: 88			
; TYPE: PRT			
; ORGANISM: Acidothermus cellulolyticus			
; FEATURE:			
; OTHER INFORMATION: Carbohydrate binding domain			
US-10-155-400-5			
Query Match 99.8%; Score 469; DB 14; Length 88;			
Best Local Similarity 100.0%; Pred. No. 3.6e-48;			
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSITVYRYWFTRDGSGSTLVYVNC	60
Db	1	VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSITVYRYWFTRDGSGSTLVYVNC	60
Qy	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
Db	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
RESULT 3			
US-09-917-376-4			
; Sequence 4, Application US/09917376			
; Publication No. US20040038334A1			
; GENERAL INFORMATION:			
; APPLICANT: DING, SHI-YOU			
; APPLICANT: ADNEY, WILLIAM S.			
; APPLICANT: VINZANT, TODD B.			
; APPLICANT: HIMMEL, MICHAEL E.			
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS			
; TITLE OF INVENTION: CELLULOXYTICUS			
; FILE REFERENCE: 40197.4US01			
; CURRENT APPLICATION NUMBER: US/09/917,376			
; CURRENT FILING DATE: 2001-07-28			
; NUMBER OF SEQ ID NOS: 7			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 4			
; LENGTH: 89			
; TYPE: PRT			
; ORGANISM: Acidothermus cellulolyticus			
; FEATURE:			
; NAME/KEY: MOD RES			
; LOCATION: (89)			
; OTHER INFORMATION: Any amino acid			
; OTHER INFORMATION: Carbohydrate binding domain			
US-09-917-376-4			
Query Match 99.8%; Score 469; DB 11; Length 89;			
Best Local Similarity 100.0%; Pred. No. 3.6e-48;			
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSITVYRYWFTRDGSGSTLVYVNC	60
Db	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
Qy	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
Db	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
RESULT 4			
US-10-155-400-4			
; Sequence 4, Application US/10155400			
; Publication No. US20030108988A1			
; GENERAL INFORMATION:			
; APPLICANT: DING, SHI-YOU			
; APPLICANT: ADNEY, WILLIAM S.			
; APPLICANT: VINZANT, TODD B.			
; APPLICANT: HIMMEL, MICHAEL E.			
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS			
; TITLE OF INVENTION: CELLULOXYTICUS			
; FILE REFERENCE: NREL 01-36A			
; CURRENT APPLICATION NUMBER: US/10/155,400			
; CURRENT FILING DATE: 2002-10-22			
; NUMBER OF SEQ ID NOS: 7			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 4			
; LENGTH: 89			
; TYPE: PRT			
; ORGANISM: Acidothermus cellulolyticus			
; FEATURE:			
; NAME/KEY: MOD RES			
; LOCATION: (89)			
; OTHER INFORMATION: Any amino acid			
; OTHER INFORMATION: Carbohydrate binding domain			
US-10-155-400-4			
Query Match 99.8%; Score 469; DB 14; Length 89;			
Best Local Similarity 100.0%; Pred. No. 3.6e-48;			
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSITVYRYWFTRDGSGSTLVYVNC	60
Db	1	VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSITVYRYWFTRDGSGSTLVYVNC	60
Qy	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
Db	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
RESULT 5			
US-09-917-378-4			
; Sequence 4, Application US/09917378			
; Publication No. US20030119093A1			
; GENERAL INFORMATION:			
; APPLICANT: DING, SHI-YOU			
; APPLICANT: ADNEY, WILLIAM S.			
; APPLICANT: VINZANT, TODD B.			
; APPLICANT: DECKER, STEPHEN R.			
; APPLICANT: HIMMEL, MICHAEL E.			
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS			
; TITLE OF INVENTION: CELLULOXYTICUS			
; FILE REFERENCE: 40197.7US01			
; CURRENT APPLICATION NUMBER: US/09/917,378			
; CURRENT FILING DATE: 2001-07-28			
; NUMBER OF SEQ ID NOS: 8			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 4			
; LENGTH: 154			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate			
; OTHER INFORMATION: binding domain			
US-09-917-378-4			
Qy	1	VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSITVYRYWFTRDGSGSTLVYVNC	60
Db	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
Qy	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88
Db	61	WAAMGCGNIRASFGSVNPATPTADTYLQ	88

OTHER INFORMATION: Guxa
US-09-917-384-5
Query Match 99.1%; Score 466; DB 10; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.6e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSSTLVYVNC 60
Db 1 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSSTLVYVNC 60
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 8
US-09-917-383-5
; Sequence 5, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-5
Query Match 99.1%; Score 466; DB 10; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.6e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSSTLVYVNC 60
Db 1 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSSTLVYVNC 60
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 9
US-09-917-384-6
; Sequence 6, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6

Query Match 99.8%; Score 469; DB 10; Length 154;
Best Local Similarity 100.0%; Pred. No. 7e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSSTLVYVNC 60
Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSSTLVYVNC 60
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
RESULT 6
US-09-917-378-1
; Sequence 1, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-917-378-1
Query Match 99.8%; Score 469; DB 10; Length 762;
Best Local Similarity 100.0%; Pred. No. 4.9e-47;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSSTLVYVNC 60
Db 455 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSSTLVYVNC 514
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 515 WAAMCGNIRASFGSVNPATPTADTYLQ 542
RESULT 7
US-09-917-384-5
; Sequence 5, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOXYLICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of


```
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-6

Query Match          99.1%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.6e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 477 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 536
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564
   |||:|||||

RESULT 10
US-09-917-383-6
; Sequence 6, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-6

Query Match          99.1%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.6e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 477 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 536
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564
   |||:|||||

RESULT 11
US-09-917-384-1
; Sequence 1, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
```

```
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-1

Query Match          99.1%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 2e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 584 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 643
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671
   |||:|||||

RESULT 12
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-1

Query Match          99.1%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 2e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 584 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 643
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671
   |||:|||||

RESULT 13
US-09-917-376-1
; Sequence 1, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
```



```
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
US-09-917-376-1

Query Match          98.3%; Score 462; DB 11; Length 957;
Best Local Similarity 97.7%; Pred. No. 4.4e-46;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYNCD 60
Db 869 VSGGVKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYNCD 928

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 14
US-10-155-400-1
; Sequence 1, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
US-10-155-400-1

Query Match          98.3%; Score 462; DB 14; Length 957;
Best Local Similarity 97.7%; Pred. No. 4.4e-46;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYNCD 60
Db 869 VSGGVKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYNCD 928

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 15
US-10-156-761-8100
; Sequence 8100, Application US/10156761
```

```
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-273697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8100
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-8100

Query Match          54.7%; Score 257; DB 14; Length 741;
Best Local Similarity 56.3%; Pred. No. 1.2e-21;
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 2 SGGVKVOYKNDSSAPGDNQIKPGLQVNTGSSVDLSTVTYVYFTRDGGSSSTLYNCDW 61
Db 591 SGGKLVLYKNDSSATDNAIKPGIRIVNTGSGSLDLKVTARYYFTRDGGSPYNAWCDY 650

QY 62 AAMCGNIRASFGSVNPATPTADTYLQ 88
Db 651 AAVGCSNVSLKVVPLTTPVPGADAYLE 677

Search completed: October 5, 2005, 08:08:32
Job time : 12.3031 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 5, 2005, 07:09:01 ; Search time 1171.63 Seconds
(without alignments)
3680.789 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVQVQYKNDSPAGDQ.....RASFGVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -US09917376/runat 04102005.164328 26897/app query.fasta 1.2524
-Q=/cgn2 1/USPTO spoel/US09917376 -SUPFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-DB=GenEmbl -OFMT=fastap -TRANS=human40.cdi -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-DOALIGN=200 -THR SCORE=pct -THR_MAX=100 -THR_LEN=2000000000
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CNC 1 1 15287 @runat 04102005.164328 26897 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Match Length	Description
1	469	99.8	2289 6 AX7000050 Sequence
2	466	99.1	3687 6 AX7000036 Sequence
3	462	98.3	2869 6 AX7000058 Sequence
4	436	92.8	3365 6 AX7000025 Sequence

5	257	54.7	299175	1	AP005023	AP005023 Streptomycin
6	228.5	48.6	1957	1	AY298814	AY298814 Thermobif
7	225.5	48.0	4567	1	AF163837	AF163837 Caldicell
8	214.5	45.6	3237	1	AF200304	AF200304 Caldicell
9	209.5	44.6	3509	1	AF488933	AF488933 Paenibac
10	204.5	43.5	4161	1	BS133614	AJ133614 Bacillus
11	195.5	41.6	2029	6	E35142	E35142 Truncated c
12	195.5	41.6	2029	6	E35143	E35143 Truncated c
13	194.5	41.4	6005	1	AF078038S5	AF078042 Caldicell
14	194.5	41.4	6416	6	E35100	E35100 Truncated c
15	192.5	41.0	4743	1	AB008029	AB008029 Thermophi
16	192.5	41.0	11707	1	AF078737	AF078737 Caldicell
17	192.5	41.0	11707	6	E35099	E35099 Truncated c
18	191.5	40.7	787	1	AF078038S3	AF078040 Caldicell
19	191	40.6	5513	1	AT286105	286105 A. thermophi
20	190.5	40.5	3262	1	AF078038S1	AF078038 Caldicell
21	190.5	40.5	5437	1	CASR69XYN2	AF036924 Caldicell
22	190	40.4	1523	1	AF355629	AF355629 Bacillus
23	190	40.4	1800	1	BSEGLSG	229076 Bacillus su
24	190	40.4	1920	1	BACGLUB	M16195 B. subtilis
25	190	40.4	2314	1	BS14GLUC	X67044 B. subtilis
26	190	40.4	2435	1	BSBGLUC2	X04689 Bacillus su
27	190	40.4	2589	1	AY044252	AY044252 Bacillus
28	190	40.4	26170	1	BC170DEGR	273234 B. subtilis
29	190	40.4	122883	1	BAM575417	AJ575417 Bacillus
30	190	40.4	122883	1	BAM576102	AJ576102 Bacillus
31	190	40.4	207829	1	BSUB0010	299113 Bacillus su
32	188	40.0	1928	1	BACCMCASE	D01057 B. subtilis
33	187	39.8	1593	1	AY183475	AY183475 Bacillus
34	187	39.8	2084	1	AF045482	AF045482 Bacillus
35	186.5	39.7	4977	1	CDCMANA	L01257 Caldocellum
36	186.5	39.7	5439	1	CDCCELA	L32742 Caldocellum
37	186	39.6	2175	1	BACCELD	M28332 B. subtilis
38	186	39.6	2175	6	E05425	E05425 DNA sequenc
39	186	39.6	2587	1	AB016164	AB016164 Bacillus
40	182.5	38.8	2977	6	A28170	A28170 B. lautus st
41	182.5	38.8	4241	1	CSCELB	X13602 Caldocellum
42	182.5	38.8	5284	1	CSU16308	U16308 Caldocellum
43	180.5	38.4	2831	1	BACCCELA	M76588 Bacillus la
44	180	38.3	1553	1	BSU27084	U27084 Bacillus sp
45	177	37.7	2152	1	AF363635	AF363635 Bacillus

ALIGNMENTS

RESULT 1
AX7000050
LOCUS AX7000050 2289 bp DNA linear PAT 03-APR-2003
DEFINITION Sequence 2 from Patent WO03012110.
ACCESSION AX7000050
VERSION AX7000050.1 GI:29536020
KEYWORDS
SOURCE
ORGANISM
Acidothermus cellulolyticus
Acidothermus cellulolyticus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Frankineae; Acidothermaceae; Acidothermus.
REFERENCE
1
AUTHORS Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.
TITLE Thermal tolerant mannanase from Acidothermus cellulolyticus
JOURNAL Patent: WO 03012110-A 2 13-FEB-2003;
Midwest Research Institute (US)
FEATURES
source
1. .2289
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN			
Alignment Scores:			
Pred. No.:	3.59e-48	Length:	2289
Score:	469.00	Matches:	88
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0


```
Query Match: 99.79% Indels: 0
DB: 6 Gaps: 0
US-09-917-376-4 (1-89) x AX700050 (1-2289)

Qy 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTGAAGTGCAGTACAAACAATGATTCGGCGCGGGTGATAACCCAG 1422

Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAAAACGGGTCTCCAGTTGGTGAATACGGGGTGTGCTCGGTGGATTTGTCACCGGTG 1482

Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGTGGCGTACTGGTTCACCGGGATGTTGGTGTGTCGACACTGGTGTACACTGTGAC 1542

Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGCGCGCGATGGGTGTGGGAATATCCGCGCTCGTTCGGCTCGGTGAACCCGCGCAG 1602

Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACCTACTCTGCAG 1626

RESULT 2
AX700036 3687 bp DNA linear PAT 03-APR-2003
LOCUS
DEFINITION Sequence 2 from Patent WO03012109.
ACCESSION AX700036
VERSION AX700036.1 GI:29536019
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. :2869
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 3.55e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 6 Gaps: 0
US-09-917-376-4 (1-89) x AX700058 (1-2869)

Qy 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAAGTGCAGTATAGAAATATGATTCGGCGCGGGTGATAATCAG 2664

Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAACCGGGTTCGACGGTGGTGAATACCGGGTGTGTCGGTGGATTTGTCACGGTG 2724

Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGGTTCAACCGCGATGGTGGCTCGTGCACACTGGTGTACAACTGTGAC 2784

Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGATGGGTGTGGGAATATCCGCGCTCGTTCGGTGGTGAACCCGCGCAG 2844

Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACCTACTCTGCAG 2868

RESULT 4
AX700025 3365 bp DNA linear PAT 03-APR-2003
LOCUS
DEFINITION Sequence 2 from Patent WO03012095.
ACCESSION AX700025
VERSION AX700025.1 GI:29536018
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
Thermal tolerant exoglucanase from Acidothermus cellulolyticus
```

```
Qy 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAAGTGCAGTACAAACAATGATTCGGCGCGGGTGATAACCCAG 1809

Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACGGGTCTCCAGTTGGTGAATACCGGGTGTGCTCGGTGGATTTGTCACCGGTG 1869

Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGTGGCGTACTGGTTCACCGGGATGTTGGTGTGTCGACACTGGTGTACAACTGTGAC 1929

Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCGATGGGTGTGGGAATATCCGCGCTCGTTCGGCTCGGTGAACCCGCGCAG 1989
```

```
Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACCTACTCTGCAG 2013

RESULT 3
AX700058 2869 bp DNA linear PAT 03-APR-2003
LOCUS
DEFINITION Sequence 2 from Patent WO03012090.
ACCESSION AX700058
VERSION AX700058.1 GI:29536021
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. :2869
/organism="Acidothermus cellulolyticus"
/mol_type="unassigned DNA"
/db_xref="taxon:28049"

ORIGIN
Alignment Scores:
Pred. No.: 3.55e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 6 Gaps: 0
US-09-917-376-4 (1-89) x AX700058 (1-2869)

Qy 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAAGTGCAGTATAGAAATATGATTCGGCGCGGGTGATAATCAG 2664

Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAACCGGGTTCGACGGTGGTGAATACCGGGTGTGTCGGTGGATTTGTCACGGTG 2724

Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGGTTCAACCGCGATGGTGGCTCGTGCACACTGGTGTACAACTGTGAC 2784

Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGATGGGTGTGGGAATATCCGCGCTCGTTCGGTGGTGAACCCGCGCAG 2844

Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACCTACTCTGCAG 2868

RESULT 4
AX700025 3365 bp DNA linear PAT 03-APR-2003
LOCUS
DEFINITION Sequence 2 from Patent WO03012095.
ACCESSION AX700025
VERSION AX700025.1 GI:29536018
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
Thermal tolerant exoglucanase from Acidothermus cellulolyticus
```


Sakaki, Y. and Hattori, M.
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2-49-10 Nishihara, Shibuya-Ku, Tokyo 151-0066, Japan
(E-mail:bio@nitech.go.jp, URL:http://www.bio.nitech.go.jp/,
Tel:81-3-3481-1933, Fax:81-3-3481-8424)
This work was done in collaboration with Haruo Ikeda(*1), Jun
Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi
Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi
Oonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi
Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7)
and Satoshi Omura(*1,*3).
Final finishing process and all annotation were done by H. Ikeda
and J. Ishikawa.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.kitasato-u.ac.jp.
Location/Qualifiers
1. .299175
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227982"
/notes="This strain is also named as strain: ATCC 31267,
NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
avermectinus"
/complement(155. .550)
/notes="SAV445
Tn4561, truncated"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC68156.1"
/db_xref="GI:29604084"
/translation="WAGPQPSRSTTTADREDSAHAPADRADEPGYRRQIKVQANLQ
EGHALARRAQLYQRYQDGMEDQIGALGLVNLVLFNTRYMDAALTQLRADGFEVR
DEDVARLUSPFMRHHVILGRYSFQLPDL"
/complement(474. .740)
/notes="IS630 family, truncated
SAV446"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC68156.1"
/db_xref="GI:29604085"
/translation="VMPDTQNAQPSSTVAAAIVTAGSVYSMMAPTSFVWAPPSPGRRR
EPGSLGAVRAYDVIRMLSRDRPTPLGDALTYYGSRRLCTSCA"
766. .987
/notes="SAV447"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="BAC68157.1"
/db_xref="GI:29604086"
/translation="MVVAVGCALTSTVPTAGGVVAEPVCVRRLTDQEGKQLQIVRRRG
SASVSRFRANMLLASAGNRVPTIAQLV"
/complement(1954. .2571)
/notes="SAV448"
/codon_start=1
/transl_table=11
/product="putative TetR-family transcriptional regulator"
/protein_id="BAC68158.1"
/db_xref="GI:29604087"
/translation="WPTSTRRANRTRRIFIEAALHVLETEGIALTIRRIATDVEVE
YSAFVYQHPANKDALVLELVAGHGLMSEFSQAQEPDTRMTIRIASQYVRFAGE
HPHLFQVWMDPMDADERRVAEPVIDVLKELTTSWAHDVVLADFDQACEILWGLT

Thu Oct 6 11:20:46 2005

```

source
1. .3509
/organism="Paenibacillus sp. BP-23"
/mol_type="genomic DNA"
/strain="BP-23"
/db_xref="taxon:198119"
/country="Spain:Catalonia, Ebro river delta"
20. .3472
/gene="cel48C"
/promoter
20. .25
/gene="cel48C"
20. .25
/gene="cel48C"
39. .45
/gene="cel48C"
62. .3337
/gene="cel48C"
/EC number="3.2.1.91"
/function="cellulose hydrolysis"
/codon_start=1
/transl_table=11
/product="cellulose 1,4-beta-cellobiosidase"
/protein_id="CAD32945.1"
/db_xref="GI:21449824"
/db_xref="GOA:O8KKF7"
/db_xref="UniProt/Tx:EMBL:O8KKF7"
/translation="MLKTAAKSLTAMLAGTVMLTGYTLGTSQTAHSDAOLAEIOA
DGNEARFLQLYALQKDPANGYFAEGIPYHSIETLUSEAPDYGHMSTSEAYSLWL
ETMYGHYTGWSKLEAAWNNMKYII PVNEGDNBEQPTTKSKNPNSPATYAAEKPF
DOYPSNLNGQYAAAGKQPLDAELKATYNNQTYLHMWLDVNDWYGYGNLLNPSTSTY
VNTFORGEQESWEAI PHSQDDKSPGKAGEGFMSTFKYKFSQAPSAQWRYPNATDADA
RAVQVLYWAKEMGYNNSEYLDKAKMGDFLRYGYMYDKYFQKVGSAKNGTPPGTGKDS
NMVLYMYTSWGGGGLGGQDWMNRIGASHTHQAYQNPAAYALSDPAGGLIPDSPTAK
ADNATLKQLEFYTLQSHHEGYAVGGATNSIGSYAAYAGVSTFDYMAVOEAPVR
DPDSTNFWFGQAWPLERVAEMYI LAESGDLTSENFMQAKKVIITKWDVTXKYVFPVG
RPVTDAAQYVLAAGQRI LGSTNAQTATPAGFEFWI PGQOEWQGPDKWNGFSTYT
NPNFHALTKDPQDVTGLGSYIKALTFFAAGTEANGTLSAKGOEAKDLAQSLDSTW
DYNDGCVITVEERKDYFRYFAKEIYLPANWSGTFQGNTPGTAGVPSDFKAGNGV
YIGYDLRPAIKQDPWAYLDNDKTKTSINTPTTKOWENGAPFTFTHRFWSQVDMATGY
EPRLLGSDSTPGVPSFAGVTAAGDQKVLSWNAAGATSYISVRAIVNGSGPT
SVATQGSFTFTDGLTNGTAYYVYI TAVNSAGESAPSAQVATPQAGVTPVPGFTLS
MGADGAAVLNWTASTGAATVTVORSTGTGAYANLATLTALTYDTTAVNGTAYNR
VIATNSAGOTNSNVVLTPMAAPI STGTLEVOYRNGSSAGNALI TPQNLKNTGTTA
IDUSKVRYFYFTKDSADMSFWCDYLAQLSANVQGSFVAVNPAKGTADTYLELSFSA
GAGSLAAGAETGIIQTFRSKNNWTFNQSDNDSYNATQTPAGATKVTAYODGAKWG
MEP"
62. .166
/gene="cel48C"
167. .3334
/gene="cel48C"
/product="cellulose 1,4-beta-cellobiosidase"
/EC number="3.2.1.91"
3456. .3472
/gene="cel48C"

sig_peptide
mat_peptide
terminator

ORIGIN
Alignment Scores:
Pred. No.: 1.35e-15 Length: 3509
Score: 209.50 Matches: 40
Percent Similarity: 70.11% Conservative: 21
Best Local Similarity: 45.98% Mismatches: 25
Query Match: 44.57% Indels: 1
DB: 1 Gaps: 1

US-09-917-376-4 (1-89) x PAE488933 (1-3509)
QY 2 SerGlyGlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 2879 ACGGACGCTTAAATCTGAAAANAATACAGGTACACGGCAATTTGATCTGAGTAAGGTGAAG 2938
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 2939 ACGCGCAATTAATCTGAAAANAATACAGGTACACGGCAATTTGATCTGAGTAAGGTGAAG 2998

```


/db_xref="taxon:32630"

ORIGIN

Alignment Scores:
Pred. No.: 3,78e-14 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.60% Indels: 1
DB: 6 Gaps: 1

US-09-917-376-4 (1-89) x E35142 (1-2029)

QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1 ATGGGAAGTGGTGAAGTACTGTACAAAGCAATGAGACGAAGTGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
DB 61 ATAGCGCGTGGTTTAAGATAGTGAATGGAGGCAGCAGCAGGTTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 121 AAGTAAGACTACTGTACACAGTGCATGGTGCACAGCCACAGAGTCCGTA---TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 178 TGGGCACAGATAGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 238 AGTGGAGCGGATTATTACCTGGAG 261

RESULT 12

E35143
LOCUS E35143 2029 bp DNA linear PAT 18-JUN-2001
DEFINITION Truncated cellulase composition.
ACCESSION E35143
VERSION E35143.1 GI:13018968
KEYWORDS JP 1999221086-A/45.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 2029)
AUTHORS Paiji,A., Peters,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,
Hyu,M. and Dajan,P.W.
TITLE Truncated cellulase composition
JOURNAL Patent: JP 1999221086-A 45 17-AUG-1999;
CLARIANT INTERNATIONAL LTD
COMMENT OS Unidentified
PN JP 1999221086-A/45
PD 17-AUG-1999
PF 21-SEP-1998 JP 1998283606
PR 19-SEP-1997 US 08/932571
PI PAIJI ANDERSON, PETAER L BAGUKUISUTO, ROY M DANIEL, PI
GURAHAMU K FARINTON,
PI MORELAND DAVID GIBUSU, HYU MORGAN, DAJAN PURATONITISU WILLIAM
PC C12N15/09, C12N15/386, C12N1/21, C12N9/42, C12N1/21, C12R1:19, PC
(C12N9/42, C12R1:19), C12N15/00
CC

Key Location/Qualifiers
FH source 1. 2029
FT source /organism='Unidentified'.
FEATURES
source
1. 2029 Location/Qualifiers
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'

ORIGIN

Alignment Scores:
Pred. No.: 3,78e-14 Length: 2029
Score: 195.50 Matches: 38

Percent Similarity: 60.23%
Best Local Similarity: 43.18%
Query Match: 41.60%
DB: 6

US-09-917-376-4 (1-89) x E35143 (1-2029)

QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1 ATGGGAAGTGGTGAAGTACTGTACAAAGCAATGAGACGAAGTGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
DB 61 ATAGCGCGTGGTTTAAGATAGTGAATGGAGGCAGCAGCAGGTTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 121 AAGTAAGACTACTGTACACAGTGCATGGTGCACAGCCACAGAGTCCGTA---TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 178 TGGGCACAGATAGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 238 AGTGGAGCGGATTATTACCTGGAG 261

RESULT 13
AF07803855
LOCUS AF07803855 6005 bp DNA linear BCT 11-FEB-2001
DEFINITION Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial
cgs: and Cels gene, complete cds.

ACCESSION AF078042
VERSION AF078042.1 GI:12743878
KEYWORDS
SEGMENT
SOURCE
ORGANISM

5 of 5
Caldicellulosiruptor sp. Tok7B.1
Caldicellulosiruptor sp. Tok7B.1
Bacteria; Firmicutes; Clostridia; Clostridiales;
Syntrophomonadaceae; Caldicellulosiruptor.

REFERENCE 1 (bases 1 to 6005)
AUTHORS

TITLE Williams,D.P. and Bergquist,P.L.
Multidomain and multifunctional glycosyl hydrolases from the
extreme thermophile caldicellulosiruptor isolate Tok7B.1
Curr. Microbiol. 40 (5), 333-340 (2000)

JOURNAL MEDLINE
PUBMED 10706665

REFERENCE 2 (bases 1 to 6005)
AUTHORS

Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Bergquist,P.L.

TITLE Direct Submission
JOURNAL

Submitted (27-OCT-1999) Biological Sciences, Macquarie University,
Sydney, NSW 2109, Australia

FEATURES
source

1. 6005 Location/Qualifiers
/organism='Caldicellulosiruptor sp. Tok7B.1'
/mol_type='genomic DNA'
/isolate='Tok7B.1'
/db_xref='taxon:80339'

CDS

<1..506
/note='ORF6; multidomain'
/codon_start=3
/transl_table=11
/product='glycosyl hydrolase 6'
/protein_id='AAK06393.1'
/db_xref='GI:12743884'

/translation='GQPDVWNGTYTGNPNLHVKIVDYGTDLGITASLANALLYSAAT
KEYGVSDEAAKLNKLELDRLWNLVDDKGLSAPKKRFEQVEYVIGAWTGM
PNGDIVKSGVKFIDIRSKYKQDFDQKLVSNAYNAGEAPEFRHFRWQCDIAIANATY
EILFGNQ'

misc_feature

<1..503
/note='Region: family 48 glycosyl hydrolase domain'

Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
| | | | | : : : : : | | | | | : : : : :
Db 1449 AGCGGACAGATAAGGTACTGTATGCTAAACAAGGACACAAACACGACGACGATA 1508
| | | | | : : : : : | | | | | : : : : :
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
| | | | | : : : : : | | | | | : : : : :
Db 1509 AGGCCATGGTTGAAGTAGTGAACACGGAAGCAGCAGCATAGATTGTGACGCGGTAAACG 1568
| | | | | : : : : : | | | | | : : : : :
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTyr 61
| | | | | : : : : : | | | | | : : : : :
Db 1569 ATAAGGTACTGTGTACACGCGGTGATGGAGACAGGACAGAGTGGGTATCA---GATTGG 1625
| | | | | : : : : : | | | | | : : : : :
Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
| | | | | : : : : : | | | | | : : : : :
Db 1626 GCACAGATAGGAGCAAGCAATGTAACATTCAAGTTTGTGAAGCTGAGCAGTAGCGTAAGT 1685
| | | | | : : : : : | | | | | : : : : :
Qy 82 ThrAlaAspThrTyrLeuGln 88
| | | | | : : : : : | | | | | : : : : :
Db 1686 GGAGCGGACTATTACTTTGGAG 1706
| | | | | : : : : : | | | | | : : : : :

Search completed: October 5, 2005, 16:33:45
Job time : 1212.63 secs

Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
| | | | | : : : : : | | | | | : : : : :
Db 2842 ATAGGGCAACCATGTGACATCAATTTGTGAAGCTTAGCAGCGGAGTGGAGCG 2901
| | | | | : : : : : | | | | | : : : : :
Qy 84 AspThrTyrLeuGln 88
| | | | | : : : : : | | | | | : : : : :
Db 2902 GATTATTACCTGGAG 2916
| | | | | : : : : : | | | | | : : : : :

RESULT 15
AB008029 4743 bp DNA linear BCT 22-OCT-1997
LOCUS Thermophilic anaerobe NA10 gene for beta-glucanase, complete cds.
DEFINITION
ACCESSION AB008029
VERSION AB008029.1 GI:2564014
KEYWORDS beta-glucanase.
SOURCE thermophilic anaerobe NA10
ORGANISM thermophilic anaerobe NA10
Bacteria.

REFERENCE
1 (sites)
Miyake,K., Machida,Y., Hattori,K. and Iijima,S.
AUTHORS Characterization of a multi-domain cellulase from an extremely
TITLES thermophilic anaerobe strain NA10
JOURNAL Unpublished
2 (bases 1 to 4743)
Miyake,K., Machida,Y. and Iijima,S.
AUTHORS Direct Submission
TITLES Submitted (09-OCT-1997) Katsuhide Miyake, Nagoya University,
JOURNAL Department of Biotechnology, School of Engineering,; Furo-cho,
Chikusa-Ku, Nagoya, Aichi 464-01, Japan
(E-mail:miyake@proc.nubio.nagoya-u.ac.jp, Tel:+81-052-789-4278,
Fax:+81-052-789-3221)

FEATURES
source
1. 4743
/organism="thermophilic anaerobe NA10"
/mol_type="genomic DNA"
/strain="NA10"
/db_xref="taxon:67756"
336..3338
/codon_start=1
/transl_table=11
/product="beta-glucanase"
/protein_id="BA22939.1"
/db_xref="GI:2564015"
/translation="MEKGLILKHFNSITPENIEIPSELLINENTYNFKADSFVNFAL
SNMTIRGHTLVWQTPDFKDSNGNTLSKDALSLKQIYDVVGRYKGVYAWD
VVNEAIDRQSGDYRRSNWYICGPEYIEKAFIWAHEADPNAKLFYNDYNTENSQKRQ
FYNLIKSLKEGVPIHGVGLQCHINLWPSISEIENTIKLFSSIPGIEIHITELDMS
FYQGSSTNYAVPRELLIQOARYKELPDFLKKYKNVTVNFWGLKDDYSWLSKNY
GKDYPLLFDENYSKYAFWSLIDPSIVPTTITLTPPEIQPLITPTPTASVTPTPT
VTPTPTVTPT
PWLKVNTGSSSIDLSRVIRYWTVDGDKAQSAVDWAQIGASNVTFKVKLSSSVS
GADYYLEIGFKSGAQLOKGTGIBQIRFNKSDMSYNNQNDMSWISQMTSYGENMK
VTAYIDGLVNGQEPGTATAPIATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
TATPATPTASTAYLDDTDNDWLYVSGNKIVDKGRPWLTVGNWFGVNTGTNVED
GWSNLSKSTLBIANRFLNRVPISEAILINWSQGIKPNINITYNPELEGKNSL
EVFDIVQTCKEVGLKIMLDISIKTDAMGHYPVWYDEKFTPEDFYKACEMITNRYK
NDDTIIADPKNEPHGKFWQDTTFAKWDNSTDINNWKYAAETCAKRILININELLIVI
EGIEAYPKDDVTWTSKSSDYSTWNGNLRGRKYPINLGYQNVVSPHYDGPVS
YQOPWFYGFPTKESLQDCWRPNWAYIMEENTAPLLIGEGSHLDGADNEKWKMYLRD
YIENHRIHRTWCFNANSQDGLGVDFPTTDEKKYSFLKPALWQDSQGRFVGLDHK
RPLSTNGKNINIITYNNNEPEVPATK"

ORIGIN
Alignment Scores:
Pred. No.: 2.59e-13 Length: 4743
Score: 192.50 Matches: 38
Percent Similarity: 63.22% Conservative: 17
Best Local Similarity: 43.68% Mismatches: 31
Query Match: 40.96% Indels: 1
DB: 1 Gaps: 1

US-09-917-376-4 (1-89) x AB008029 (1-4743)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 5, 2005, 06:17:10 ; Search time 136.919 Seconds
(without alignments)
3847.935 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQVKNDSAPGDNO.....RASFGSNPATPTADTYLQX 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US0917376/runat_04102005_164327_26889/app.query.fasta_1.2524
-DB=N_Geneseq_16Dec04 -QWMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0917376 @CGN 1 1867 @runat_04102005_164327_26889 -NCPUP=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:
1: geneseq1980s:
2: geneseq1990s:
3: geneseq2000s:
4: geneseq2001as:
5: geneseq2001bs:
6: geneseq2002as:
7: geneseq2002bs:
8: geneseq2003as:
9: geneseq2003bs:
10: geneseq2003cs:
11: geneseq2003ds:
12: geneseq2004as:
13: geneseq2004bs:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	469	99.8	2289	8 ABZ77633 Nucleotid
2	469	99.8	2289	10 ADJ38292 A. cellu
3	466	99.1	3687	8 ABZ77634 Nucleotid
4	466	99.1	3687	12 ADH36637 DNA encod
5	462	98.3	2869	10 ADD22922 Acidother

6	462	98.3	2869	10	ABZ77632	Abz77632 Nucleotid
7	462	98.3	2869	12	ADO52314	Ado52314 Acidother
8	436	92.8	3365	8	ABZ76162	Abz76162 A. cellu
9	409	87.0	3364	10	ADF75896	Adf75896 Acidother
10	231.5	49.3	2600	2	AAQ15178	Aaq15178 Portion o
11	195.5	41.6	2029	2	AAZ55660	Aaz55660 DNA seque
12	195.5	41.6	2029	6	AAZ26568	Aaz26568 Active ce
13	194.5	41.4	6415	2	AAZ55662	Aaz55662 DNA seque
14	194.5	41.4	6416	6	AAZ26526	Aaz26526 Active ce
15	192.5	41.0	11706	2	AAZ55661	Aaz55661 DNA seque
16	192.5	41.0	11707	6	AAZ26525	Aaz26525 Active ce
17	190	40.4	1527	13	ADT48400	Adt48400 Bacterial
18	190	40.4	2177	11	ADO55906	Ado55906 Bacillus
19	188	40.0	1434	6	AAZ41028	Aaz41028 CMCase ge
20	188	40.0	1488	6	AAZ41028	Aaz41028 CMCase ge
21	188	40.0	2510	6	ABK53202	Abk53202 Bacillus
22	186	39.6	2175	2	AAQ49820	Aaq49820 NK-1 cell
23	182.5	38.8	2977	2	AAQ13001	Aaq13001 Endol gen
24	165	35.1	501	10	ADG14263	Adg14263 Cellulose
25	165	35.1	717	10	ADG14257	Adg14257 Cellulose
26	165	35.1	1438	3	AAZ45336	Aaz45336 DNA encod
27	165	35.1	1482	2	AAZ90978	Aaz90978 DNA encod
28	165	35.1	1482	2	AAZ31562	Aaz31562 Pectate l
29	165	35.1	5562	2	AAT86625	Aat86625 C. thermo
30	163	34.7	1500	12	ADJ35111	Adj35111 DNA encod
31	157	33.4	1314	6	ABK73393	Abk73393 Bacillus
32	157	33.4	2602	12	ADG32259	Adg32259 DNA encod
33	155.5	33.1	4059	5	AAH75059	Aah75059 Nucleotid
34	150.5	32.0	4056	5	AAH75060	Aah75060 Nucleotid
35	130.5	27.8	4452	8	ACA27429	Aca27429 Prokaryot
36	126.5	26.9	1303	10	ADC27473	Adc27473 Fusion pr
37	126.5	26.9	1747	10	ADC27474	Adc27474 Fusion pr
38	126.5	26.9	3489	10	ADC27475	Adc27475 Fusion pr
39	120.5	25.6	1775	2	AAQ13003	Aaq13003 Endo3 gen
40	114	24.3	486	2	AAQ72917	Aaq72917 Cellulose
41	114	24.3	486	2	AAV74072	Aav74072 C. cellu
42	114	24.3	499	2	AAZ24930	Aaz24930 Clostridi
43	114	24.3	507	5	AAQ11042	Aaq11042 Clostridi
44	114	24.3	768	4	AAZ86248	Aaz86248 DNA seque
45	114	24.3	768	6	ABK52403	Abk52403 E. coli c

ALIGNMENTS

RESULT 1
ABZ77633
ID ABZ77633 standard; DNA; 2289 BP.
XX
AC ABZ77633;
XX
DT 03-JUN-2003 (first entry)
XX
DE Nucleotide sequence of the ManA polypeptide.
XX
KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX food; feed; paper pulp; biofuel; manase; gene; ss.
XX
OS Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
FT CDS 1..2289
FT /tag= a
FT /product= "ManA"
XX
XX WO2003012110-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023819.
XX
PR 28-JUL-2001; 2001WO-US023819.
XX
XX (MIDE) MIDWEST RES INST.
XX
XX

XX 03-JUN-2003 (first entry)
 XX Nucleotide sequence of the GuxA polypeptide.
 XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
 XX detergent; pulp processing; paper processing; feed processing; textile;
 XX gene; ss.
 XX Acidotherrmus cellulolyticus.
 XX Key Location/Qualifiers
 XX CDS 1..3687
 XX /tag= a
 XX /product= "GuxA"
 XX WO2003012109-A1.
 XX 13-FEB-2003.
 XX 28-JUL-2001; 2001WO-US023817.
 XX 28-JUL-2001; 2001WO-US023817.
 XX (MIDE) MIDWEST RES INST.
 XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
 XX WPI: 2003-239526/23.
 XX P-PSDB; ABP73029.
 XX Novel thermal tolerant GuxA polypeptide derived from Acidotherrmus
 XX cellulolyticus, useful for reducing cellulose in a starting material, and
 XX for the conversion of biomass to biofuels and biofuel additives.
 XX Example 1; Page 23-24; 47pp; English.
 XX The present sequence encodes a GuxA polypeptide. GuxA is thermostable
 XX cellulase, and is a member of the glycoside hydrolase family of enzymes.
 XX GuxA is useful for reducing cellulose in a starting material such as
 XX agricultural biomass to sugars. This is useful in biofuel production.
 XX GuxA is also useful in the conversion of biomass to biofuels and biofuel
 XX additives, in detergents, pulp and paper processing, food and feed
 XX processing, and in textile processes. GuxA is also useful for raising
 XX polyclonal and monoclonal antibodies that are useful in purifying GuxA,
 XX or detecting GuxA polypeptide expression, as well as reagent tools for
 XX characterizing the molecular actions of GuxA polypeptides
 XX Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,33e-47 Length: 3687
 Score: 466.00 Matches: 87
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 98.86% Mismatches: 0
 Query Match: 99.15% Indels: 0
 DB: Gaps: 0
 US-09-917-376-4 (1-89) x ABZ77634 (1-3687)
 QY 1 ValSerGlyValIysValGlnTyrIysAsnAspSerAlaProGlyAspAenGln 20
 Db 1750 GTGTCGGTGGGCTCAAGTGCAGTACAGAACATGATTCGGCGCGGTGATAACACAG 1809
 QY 21 IleLysProGlyIueGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
 Db 1810 ATCAAAACCGGCTCTCAGTTGGTGAATACCGGGTCTGTCGGTGGATTGTGACGGTG 1869
 QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTrpAsnCysAsp 60
 Db 1870 ACGTGCCTGACTGGTTTACCCGGGATGGTGGTGTGCACACTGGTGTACACTGTGAC 1929
 QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80

Db 1930 TGGCGCGGATGGGGTGTGGGAATATCCGCGCTCGTTCGGTCCGTGAACCGGCGACG 1989
 QY 81 ProThrAlaAspThrTyrLeuGln 88
 Db 1990 CCGACGGCGGACACCTACTGTCAG 2013
 RESULT 4
 ID ADH36637 standard; DNA; 3687 BP.
 XX ADH36637;
 XX 11-MAR-2004 (first entry)
 XX DNA encoding Acidotherrmus cellulolyticus glycoside hydrolase, GuxA.
 XX Thermal tolerant cellulase; glycoside hydrolase; GuxA;
 XX carbohydrate degradation; cellulose; agricultural biomass;
 XX municipal solid waste; thermostable; gene; ds.
 XX Acidotherrmus cellulolyticus.
 XX Key Location/Qualifiers
 XX CDS 1..3687
 XX /tag= a
 XX /product= "GuxA"
 XX US2003104522-A1.
 XX 05-JUN-2003.
 XX 28-JUL-2001; 2001US-00917383.
 XX 28-JUL-2001; 2001US-00917383.
 XX (DING/) DING S.
 XX (ADNE/) ADNEY W S.
 XX (VINZ/) VINZANT T B.
 XX (HIMM/) HIMMEL M E.
 XX (DECK/) DECKER S R.
 XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
 XX WPI: 2004-106451/11.
 XX P-PSDB; ADH36636.
 XX Novel isolated thermostable GuxA polypeptide useful for detecting
 XX polynucleotide encoding GuxA, assessing carbohydrate degradation activity
 XX of GuxA, reducing cellulose in starting material e.g., agricultural
 XX biomass.
 XX Claim 21; SEQ ID NO 2; 20pp; English.
 XX The present invention relates to the isolation of a thermal tolerant
 XX cellulase from Acidotherrmus cellulolyticus. The cellulase is a member of
 XX the glycoside hydrolase family and is designated GuxA. Also disclosed are
 XX methods of producing and using GuxA. The GuxA polypeptide is useful for
 XX the detection of a polynucleotide encoding GuxA. The polypeptide sequence
 XX is also useful for assessing the carbohydrate, e.g. cellulose,
 XX degradation activity of GuxA. GuxA is useful for reducing cellulose in a
 XX starting material such as agricultural biomass or municipal solid waste.
 XX The polypeptide molecule of GuxA is thermostable. The present sequence
 XX encodes A. cellulolyticus glycoside hydrolase, GuxA.
 XX Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,33e-47 Length: 3687
 Score: 466.00 Matches: 87
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 98.86% Mismatches: 0
 Query Match: 99.15% Indels: 0


```

DB: 12 Gaps: 0
US-09-917-376-4 (1-89) x ADH36637 (1-3687)
QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTGGGTGGCTGAGGTGACGTACAGAACATGATTCCGGCCCGGTGATAACACAG 1809
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAACCGGGTCTCCAGTGGTGAATACCGGGTCTGTCGTGGATTTGTCCACGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGGGTACTGTTCCACCGGGATGTTGGTCTGTCGACACTGGTGTCAACACTGTGAC 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGGATGGGGTGTGGGAATATCCGCGCTCGTTCGGCTCGTGAACCCGCGACG 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGACACCTACCTGCAG 2013
RESULT 5
ADD22922
ID ADD22922 standard; DNA; 2869 BP.
XX AC ADD22922;
XX DT 15-JAN-2004 (first entry)
XX DE Acidothermus cellulolyticus avicelase AvIII DNA.
XX KW AvIII; cellulose reduction; agricultural biomass; municipal solid waste;
XX KW glycoside hydrolase; avicelase; ds; gene.
XX OS Acidothermus cellulolyticus.
XX FH Key Location/Qualifiers
XX FT CDS 1..2868
XX FT /*tag= a
XX FT /product= "AvIII"
XX PN US2003108988-A1.
XX PD 12-JUN-2003.
XX PF 18-OCT-2002; 2002US-00155400.
XX PR 28-JUL-2001; 2001US-00917376.
XX (DING/) DING S.
XX PA (ADNE/) ADNEY W S.
XX PA (VINZ/) VINZANT T B.
XX PA (HIMM/) HIMMEL M E.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-810853/76.
XX DR P-PSDB; ADD22921.
XX New isolated thermal tolerant avicelase polynucleotide useful for
XX detection of a polynucleotide encoding AvIII and for reducing cellulose
XX in a starting material, e.g. municipal solid waste.
XX Claim 17; SEQ ID NO 2; 29pp; English.
XX The invention relates to an isolated polynucleotide molecule encoding a
XX thermostable AvIII polypeptide. The polynucleotide is useful for
XX detection of a polynucleotide encoding AvIII. The polynucleotide is
XX useful for reducing cellulose in a starting material which involves
XX administering to the starting material, e.g. agricultural biomass or

```

```

CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents DNA encoding the Acidothermus cellulolyticus
CC avicelase AvIII.
XX SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;
Alignment Scores: 5.3e-47 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: Gaps: 0
US-09-917-376-4 (1-89) x ADD22922 (1-2869)
QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTGGGTGGGTGAGGTGACGTATAGAAATATGATTCCGGCCCGGTGATAATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAACCGGGTGTGACGTGGTGAATACCGGGTCTGTCGTGGATTTGTCCACGGTG 2724
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGGGTACTGTTCCACCGGGATGTTGGTCTGTCGACACTGGTGTCAACACTGTGAC 2784
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGGATGGGGTGTGGGAATATCCGCGCTCGTTCGGCTCGTGAACCCGCGACG 2844
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGACACCTACCTGCAG 2868
RESULT 6
ABZ77632
ID ABZ77632 standard; DNA; 2869 BP.
XX AC ABZ77632;
XX DT 03-JUN-2003 (first entry)
XX DE Nucleotide sequence of the avicelase AvIII.
XX KW Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
XX KW detergent; pulp processing; paper processing; feed processing; textile;
XX KW cellulose; gene; ss.
XX OS Acidothermus cellulolyticus.
XX FH Key Location/Qualifiers
XX FT CDS 1..2869
XX FT /*tag= a
XX FT /partial
XX FT /product= "AvIII"
XX FT /transl_except= (pos:2869,aa:Xaa)
XX FT /note= "Xaa is an unspecified residue"
XX PN WO2003012090-A2.
XX PD 13-FEB-2003.
XX PF 28-JUL-2001; 2001WO-US023818.
XX PR 28-JUL-2001; 2001WO-US023818.
XX PA (MIDE ) MIDWEST RES INST.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;

```


XX WPI; 2003-248177/24.
 DR P-PSDB; ABP73015.
 XX
 XX New thermostable AvIII peptide from Acidothermus cellulolyticus, useful
 PT for degradation of cellulose or in generating anti-AvIII antibodies for
 PT purifying recombinant AvIII polypeptides from genetically engineered
 PT host cells.
 XX
 XX Claim 3; Page 24; 44pp; English.
 PS
 XX The present sequence encodes a thermostable avicelase polypeptide,
 CC designated AvIII. AvIII is a member of the glycoside hydrolase family
 CC of enzymes, and is a cellulase. AvIII is useful in the conversion of
 CC biomass to biofuels and biofuel additives. It may be useful in the
 CC production of detergents, pulp and paper processing, food and feed
 CC processing and in textile processes. The thermostable AvIII peptide is
 CC useful in the degradation of cellulose, and in generating specific anti-
 CC AvIII antibodies that are useful in purifying recombinant AvIII
 CC polypeptides from genetically engineered host cells, in detecting AvIII
 CC polypeptide expression, as well as a reagent tool for characterizing the
 CC molecular actions of the polypeptide. The AvIII polynucleotide is useful
 CC as a source of probes or primers in various diagnostic assays
 XX
 XX Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;
 SQ
 Alignment Scores:
 Pred. No.: 5.3e-47 Length: 2869
 Score: 462.00 Matches: 86
 Percent Similarity: 100.00% Conservative: 2
 Best Local Similarity: 97.73% Mismatches: 0
 Query Match: 98.30% Indels: 0
 DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x ABZ77632 (1-2869)

QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
 DB 2605 GTGTCGGTGGGGTGAAGTGCAGTATAGAAATATGATTCGGCGCCGGTGATATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
 DB 2665 ATCAAGCCGGGTTCAGGTGTGAATACCGGGTCGTGGTGATTTGTCGACGGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
 DB 2725 ACGTGGCGTACTGGTTTACCCGGGATGGTGGCTCGTCGACACTGGTGACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
 DB 2785 TGGCGCGGATCGGGTGTGGGAATATCCGCGCTCGTTCGGTCTCGTGAAACCGCGCAGC 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
 DB 2845 CCGACGGCGGACACCTACCTGCAG 2868

RESULT 7
 ADO52314
 ID ADO52314 standard; DNA; 2869 BP.
 XX
 XX ADO52314;
 AC
 XX 15-JUL-2004 (first entry)
 DT
 XX Acidothermus cellulolyticus avicelase III (AvIII) DNA.
 DE
 XX Thermostable cellulase; avicelase III; AvIII; cellulose reduction;
 KW biomass degradation; ethanol formation; industrial chemical;
 KW fabric treatment; gene; ds.
 KW
 XX Acidothermus cellulolyticus.
 OS
 XX Location/Qualifiers
 FH

CDS
 FT 1..2868
 FT /*tag= b
 FT /product= "Acidothermus cellulolyticus avicelase III"
 FT /note= "CDS contains translational exceptions"
 FT
 FT /partial
 FT /note= "No stop codon"
 FT 1..306
 FT sig_peptide /*tag= a
 FT 307..2865
 FT mat_peptide /*tag= c
 FT
 FT /product= "Acidothermus cellulolyticus mature avicelase
 FT III"
 FT 2869
 FT misc_feature /*tag= d
 FT /note= "A string of unknown nucleic acid units"
 FT
 XX US2004038334-A1.
 XX 26-FEB-2004.
 XX 28-JUL-2001; 2001US-00917376.
 XX 28-JUL-2001; 2001US-00917376.
 XX (DING/) DING S.
 XX (ADNE/) ADNEY W S.
 XX (VINZ/) VINZANT T B.
 XX (HIMM/) HIMMEL M E.
 XX
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;
 PI WPI; 2004-203224/19.
 DR P-PSDB; ADO52313.
 DR
 XX Novel thermostable AvIII polypeptide of glycoside hydrolase family and
 PT isolated from Acidothermus cellulolyticus, useful for degrading treated
 PT biomass into simpler forms of carbohydrate.
 XX
 XX Claim 17; SEQ ID NO 2; 19pp; English.
 PS
 CC The invention relates to a thermostable cellulase enzyme, avicelase III
 CC (AvIII) and its nucleic acid sequence. AvIII is useful for reducing
 CC cellulose in a starting material. A thermostable AvIII peptide is useful
 CC for degrading treated biomass into simpler forms of carbohydrate, which
 CC is used in the formation of ethanol or other industrial chemicals. It is
 CC also useful for treating fabrics to remove cellulose-containing stains.
 CC The present sequence is Acidothermus cellulolyticus AvIII DNA.
 XX
 XX Sequence 2869 BP; 546 A; 857 C; 899 G; 565 T; 0 U; 2 Other;
 SQ
 Alignment Scores:
 Pred. No.: 5.3e-47 Length: 2869
 Score: 462.00 Matches: 86
 Percent Similarity: 100.00% Conservative: 2
 Best Local Similarity: 97.73% Mismatches: 0
 Query Match: 98.30% Indels: 0
 DB: 12 Gaps: 0

US-09-917-376-4 (1-89) x ADO52314 (1-2869)

QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
 DB 2605 GTGTCGGTGGGGTGAAGTGCAGTATAGAAATATGATTCGGCGCCGGTGATATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
 DB 2665 ATCAAGCCGGGTTCAGGTGTGAATACCGGGTCGTGGTGATTTGTCGACGGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
 DB 2725 ACGTGGCGTACTGGTTTACCCGGGATGGTGGCTCGTCGACACTGGTGACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80

Db 2785 TGGGCGCGATCGGTGGGATATCGCGCTCGGTGGAACCCGCGACG 2844
 Qy 81 ProThrAlaAspThrTyrLeuGln 88
 Db 2845 CCGACGCGCGACACCTACCTGAG 2868
 RESULT 8
 ABZ76162
 ID ABZ76162 standard; DNA; 3365 BP.
 XX
 AC ABZ76162;
 DT 29-MAY-2003 (first entry)
 XX
 DE A. cellulolyticus Guxl protein encoding DNA.
 XX
 KW Guxl; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;
 biofuel; detergent; pulp; paper processing; feed processing; textile;
 cellulase; gene; ds.
 XX
 OS Acidothermus cellulolyticus.
 XX
 FH Key Location/Qualifiers
 CDS 1..3365
 FT /tag= a
 FT /product= "Guxl"
 FT /transl_except= (pos: 682..683, aa: Pro)
 FT /note= "this codon has an apparent one nucleotide
 basepair deletion which alters the reading frame"
 XX
 PN WO2003012095-A1.
 XX
 PD 13-FEB-2003.
 XX
 PF 28-JUL-2001; 2001WO-US023820.
 XX
 PR 28-JUL-2001; 2001WO-US023820.
 XX
 PA (MIDE) MIDWEST RES INST.
 XX
 PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;
 PI Lantz McCarter S;
 PI
 DR WPI; 2003-300494/29.
 DR P-PSDB; ABP71656.
 XX
 PT New thermal tolerant Guxl peptide having specified amino acid sequence,
 PT useful in the degradation of cellulose to biofuels.
 XX
 PS Disclosure; Page 22-23; 44pp; English.
 XX
 CC The invention relates to a thermal tolerant Guxl peptide from A.
 CC cellulolyticus. The Guxl exoglucanase is a member of the glycoside
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate
 CC binding domain type III, and a carbohydrate binding domain type II. The
 CC polypeptide is useful in the degradation of cellulose into biofuel, or
 CC for conversion of biomass to biofuel additives. It is used in detergents,
 CC pulp and paper processing, food and feed processing, and in textile
 CC processing. It can also be used alone or in combination with other
 CC cellulase or glycoside hydrolases. The novel polypeptide generates
 CC alternative cellulase enzymes capable of assisting in the commercial-
 CC scale processing of cellulose to sugar for use in biofuel production. The
 CC present sequence represents a A. cellulolyticus Guxl cellulase encoding
 CC DNA
 XX
 SQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.11e-43 Length: 3365
 Score: 436.00 Matches: 81
 Percent Similarity: 97.62% Conservative: 1
 Best Local Similarity: 96.43% Mismatches: 2

Query Match: 92.77% Indels: 0
 DB: 8 Gaps: 0
 US-09-917-376-4 (1-89) x ABZ76162 (1-3365)
 Qy 5 VallysValGlnTyrTlyeAsnAsnAspSerAlaProGlyAspAsnGlnIleIysProGly 24
 Db 112 CTCAAAGCGCAGTATAGAACAATGATTCGGCGCGAGTGACAACAGATCAACCGGGT 171
 Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44
 Db 172 CTCAGTTGGTGAATACCGGGTCGTCGGTGGGATTTGTCACCGTGACGGTGCCTAC 231
 Qy 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCyAspTrpAlaAlaMet 64
 Db 232 TGGTTCACCGCGGATGGTGGGTGCGTGCACACTGGTGTAACAATGTCGACCTGGCGGCGATG 291
 Qy 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84
 Db 292 GGGTGTGGGAATATCCGGCCCTCGTTCCGCTCGGTGAACCCCGCGACGCCGCGCGAC 351
 Qy 85 ThrTyrLeuGln 88
 Db 352 ACCTACCTGCAG 363
 RESULT 9
 ADF75896
 ID ADF75896 standard; DNA; 3364 BP.
 XX
 AC ADF75896;
 XX
 DT 26-FEB-2004 (first entry)
 XX
 DE Acidothermus cellulolyticus Guxl DNA.
 XX
 KW Guxl; gene; ds; thermal tolerant; cellulase; glycoside hydrolase;
 exoglucanase; cellulose; sugar; biofuel production.
 XX
 OS Acidothermus cellulolyticus.
 XX
 FH Key Location/Qualifiers
 CDS 1..3364
 FT /tag= a
 FT /product= "Guxl protein"
 FT /note= "This polynucleotide sequence contains translation
 exceptions"
 XX
 PN US2003096342-A1.
 XX
 PD 22-MAY-2003.
 XX
 PF 28-JUL-2001; 2001US-00917384.
 XX
 PR 28-JUL-2001; 2001US-00917384.
 XX
 PA (ADNE/) ADNEY W S.
 PA (DING/) DING S. T B.
 PA (VINZ/) VINZANT T B.
 PA (HIMM/) HIMMEL M E.
 PA (DECK/) DECKER S R.
 PA (MCCA/) MCCARTER S L.
 XX
 PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR, McCarter SL;
 XX WPI; 2003-863404/80.
 DR P-PSDB; ADF75895.
 XX
 PT Novel thermal tolerant cellulase of glycoside hydrolase family,
 PT comprising catalytic domain, first and second carbohydrate binding
 PT domain, isolated from Acidothermus cellulolyticus, useful in cellulose
 PT degradation.
 XX
 PS Claim 21; SEQ ID NO 2; 20pp; English.

Db 121 AAGTAAAGTACTGGTACACAGTGGTGTGACAAAGCCACAGAGTGGCGTA---TGTGAC 177
Qy 61 TrpAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 178 TGGGCACAGATAGGGCAAGCAATGTGCATTCAATTTGTGAAGCTTTCAGCGGAGTG 237
Qy 81 ProThrAlaSerThrTyrLeuGln 88
Db 238 AGTGGAGCGGATTATTACCTGGAG 261
RESULT 13
AA55662
ID AA55662 standard; DNA; 6415 BP.
AC AA55662;
DT 30-JUL-1999 (first entry)
XX DNA sequence encoding truncated cellulases.
DE Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing; ss.
XX Unidentified.
OS
XX
XX EP921188-A2.
XX
XX 09-JUN-1999.
XX
XX 15-SEP-1998; 98EP-00810919.
XX
XX 19-SEP-1997; 97US-00932571.
XX (CLRN) CLARIANT FINANCE BVI LTD.
XX
XX Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;
PI Morgan H, Williams DP;
PI
DR WPI: 1999-315403/27.
XX P-PSDB; AAY13493.
XX
XX New truncated cellulase proteins, useful in detergents and for producing
PT 'stonewashed' denim.
XX
XX Disclosure; Page 24-25; 65pp; English.
XX
XX The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from
CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3
CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751
CC and the stability region extends from amino acid E482 to G635 in the
CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or remove
CC staining, backstaining or graying, for use on cellulosic materials
CC including cotton-containing fabrics. They are especially useful for
CC preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using non-
XX truncated cellulase compositions
SQ Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;

Alignment Scores:
Pred. No.: 2,66e-13 Length: 6415
Score: 194.50 Matches: 38
Percent Similarity: 61.18% Conservative: 14

Best Local Similarity: 44.71% Mismatches: 32
Query Match: 41.38% Indels: 1
DB: 2 Gaps: 1
US-09-917-376-4 (1-89) x AA55662 (1-6415)
Qy 4 GlyValysValGlnTyrIlyAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2664 GGTGTGAAGGTACTGTGTACAAACAATGAGACAGTGGCAGCACAGGTTCTTATAAGGCCG 2723
Qy 24 GlyLeuGlnLeuValAsnThrClySerSerSerValAspLeuSerThrValThrValArg 43
Db 2724 TGGTTTAAGTAGTGAATGGAGGCAGCAGCAGTGTGTATCTTAGCAGGGTTAAGATAGA 2783
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63
Db 2784 TACTGTACACAGTGTGATGGTGCACAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 2840
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2841 ATAGGGCAAGCAATGTGCATTCAATTTGTGAAGCTTAGCAGCGGAGTGGAGCGC 2900
Qy 84 AspThrTyrLeuGln 88
Db 2901 GATTATTACCTGGAG 2915
RESULT 14
AAD26526
ID AAD26526 standard; DNA; 6416 BP.
AC AAD26526;
DT 26-MAR-2002 (first entry)
XX
XX Active cellulase protein, cele gene.
KW Active cellulase protein; alkaliphilic; textile processing; proteinase;
KW detergent additive; stonewashed appearance; cotton-containing denim;
KW CelB5; thermophilic; commercial detergent; cele gene; ds.
XX
XX Unidentified.
XX
XX Location/Qualifiers
FH Key 634..5889
FT CDS /tag= a
FT /product= "Cele protein"
FT misc_feature 748..2538
FT /tag= c
FT /product= "DNA encoding E1/2 protein"
FT misc_feature 748..2076
FT /tag= b
FT /product= "DNA encoding E1 protein"
XX
XX US6294366-B1.
XX
XX 25-SEP-2001.
XX
XX 19-AUG-1998; 98US-00136574.
XX
XX 19-SEP-1997; 97US-00932571.
XX (CLRN) CLARIANT FINANCE BVI LTD.
XX
XX Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
PI Morgan H, Williams DP;
PI
XX WPI: 2002-081780/11.
DR P-PSDB; AAE16324.
XX
XX New cellulase active protein, useful in textile processing or commercial
PT detergents, e.g. for improving the feel or appearance of cotton-
PT containing fabrics, is stable under conditions of alkaline pH and
PT elevated temperatures.

XX Disclosure; Col 37-44; 61pp; English.

XX The present invention relates to a cellulase active protein, which is

XX substantially free of proteinases of native thermophilic and

CC alkalophilic origin, where the cellulase active protein consists of the

CC CelB5 amino acid sequence. The cellulase active protein is useful for

CC treating cellulosic materials including cotton-containing fabrics, as

CC detergent additives. The cellulase active protein is also useful for

CC improving the feel and/or appearance of cotton-containing fabrics, for

CC removing surface fibers from cotton-containing knits or for imparting

CC stonewashed appearance to cotton-containing denims. The present proteins

CC are stable under condition of alkaline pH and elevated temperatures, thus

CC suitable for textile processing and in commercial detergents. The present

CC sequence is celE gene

XX Sequence 6416 BP; 2068 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;

Alignment Scores:

Pred. No.:	2.66e-13	Length:	6416
Score:	194.50	Matches:	38
Percent Similarity:	61.18%	Conservative:	14
Best Local Similarity:	44.71%	Mismatches:	32
Query Match:	41.38%	Indels:	1
DB:	6	Gaps:	1

US-09-917-376-4 (1-89) x AAD26526 (1-6416)

Qy 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23

Db 2665 GGTGTAAGGTACTGTACAGCAACATGAGCAAGTGCAGCACAGGTTCTATAAGGCG 2724

Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43

Db 2725 TGGTTTAAGTAGTAGTGAATGGAGGAGCAGCAGCAGTGTGATCTTAGCGGTTAAGATAAGA 2784

Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTIPalaAa 63

Db 2785 TACTGGTACACAGTGGATGGTGAACAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 2841

Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83

Db 2842 ATAGGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTGCAGTGGAGCG 2901

Qy 84 AspThrTyrLeuGln 88

Db 2902 GATTATTACTGGAG 2916

RESULT 15

AX55661

ID AX55661 standard; DNA; 11706 BP.

XX AC AX55661;

XX 30-JUL-1999 (first entry)

XX DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.

XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

XX Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;

XX cotton-containing fabric; stonewashing; 88.

XX Unidentified.

XX EP921188-A2.

XX 09-JUN-1999.

XX 15-SEP-1998; 98EP-00810919.

XX 19-SEP-1997; 97US-00932571.

XX (CLRN) CLARIANT FINANCE BVI LTD.

XX Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;

PI Morgan H, Williams DP;

XX WPI; 1999-315403/27.

DR P-PSDB; AAY13492.

XX New truncated cellulase proteins, useful in detergents and for producing

PT 'stonewashed' denim.

XX Disclosure; Page 20-23; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of

XX proteinases of native thermophilic and alkalophilic origin, comprising

CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,

CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-

CC length sequences, or functional equivalents. Cel B5 extends from amino

CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino

CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from

CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3

CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751

CC and the stability region extends from amino acid B482 to G635 in the

CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new

CC enzymes are useful in laundry detergent compositions to prevent or remove

CC staining, backstaining or graying, for use on cellulosic materials

CC including cotton-containing fabrics. They are especially useful for

CC preventing redeposition of colorant during stonewashing, and for

CC processing of textiles where cellulose breakdown is required. The new

CC truncated enzymes show reduced redeposition of dye compared to using non-

CC truncated cellulase compositions

XX Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.02e-12	Length:	11706
Score:	192.50	Matches:	37
Percent Similarity:	62.35%	Conservative:	16
Best Local Similarity:	43.53%	Mismatches:	31
Query Match:	40.96%	Indels:	1
DB:	2	Gaps:	1

US-09-917-376-4 (1-89) x AAX55661 (1-11706)

Qy 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23

Db 4038 GGTGTAAGGTACTGTACAGCAACATGAGCAAGTGCAGCACAGGTTCTATAAGGCG 4097

Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43

Db 4098 TGGTTTAAGTAGTAGTGAATGGAGGAGCAGCAGCAGTGTGATCTTAGCGGTTAAGATAAGA 4157

Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAa 63

Db 4158 TACTGGTACACAGTGGATGGTGAACAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 4214

Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83

Db 4215 ATAGGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTGCAGGCG 4274

Qy 84 AspThrTyrLeuGln 88

Db 4275 GATTATTACTGGAG 4289

Search completed: October 5, 2005, 09:20:53

Job time : 145.919 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 07:43:21 ; Search time 855.806 Seconds
(without alignments)
3958.515 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQYKXNDSPGDNQ.....RASFGSYNPATPTADTVLQX 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool/US09917376/runat_04102005_164328_26911/app_query.fasta_1.2524
-DB=EST -QFMT=fastap -SUPFIX=p2n.rst -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CNC 1.1.12009 @runat_04102005_164328_26911 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hcc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	81	17.2	584	CF480372	CF480372 POL1_65_H
2	79.5	16.9	754	BQ047524	BQ047524 EST596642
3	78.5	16.7	768	CB293085	CB293085 UCRCS01_0
4	77	16.4	906	BQ722991	BQ722991 AGENCOURT
5	76	16.2	376	AA043716	AA043716 zk61f10.s
6	76	16.2	606	BE362891	BE362891 DGI_90_C1
7	76	16.2	841	CD226529	CD226529 CCI_45_E
8	76	16.2	657	BE362820	BE362820 DGI_89_C0
9	76	16.2	676	CF487649	CF487649 POL1_45_C

10	76	16.2	691	4	BG739392
11	76	16.2	724	8	BZ616323
12	76	16.2	814	4	B1218457
13	76	16.2	905	8	BZ545807
14	76	16.2	934	9	CG270741
15	76	16.2	1001	9	CG270748
16	75.5	16.1	657	2	BFS13128
17	75.5	16.1	1041	9	CK1974250
18	75.5	16.1	1058	7	CK231562
19	75.5	16.1	2091	3	HSM800308
20	75	16.0	706	7	CK352071
21	74.5	15.9	465	2	AW491464
22	74.5	15.9	613	6	CA595709
23	74.5	15.9	650	8	BZ894843
24	74.5	15.7	302	7	CF076729
25	74	15.7	463	9	BX959940
26	74	15.7	553	4	B1664037
27	73.5	15.6	261	2	BF875908
28	73.5	15.6	752	7	CNI187748
29	73.5	15.6	798	7	CNI187749
30	73.5	15.6	816	6	CB293980
31	73.5	15.6	819	6	CB293084
32	73.5	15.6	894	7	CK338508
33	73.5	15.6	906	3	AY539888
34	73.5	15.6	1039	4	BG334904
35	73.5	15.6	1237	5	BQ065850
36	73	15.5	357	7	CF078232
37	73	15.5	360	6	C64738
38	73	15.5	593	7	CF622108
39	73	15.5	776	9	BX209426
40	73	15.5	861	8	BH319803
41	73	15.5	894	9	CG077853
42	73	15.5	924	9	CG077856
43	72.5	15.4	380	4	BG816375
44	72.5	15.4	450	1	AL799805
45	72.5	15.4	533	2	AW963120

ALIGNMENTS

RESULT 1
LOCUS CF480372
DEFINITION POL1_65_H07_b1_A002 Pollen Sorghum bicolor cDNA clone
ACCESSION CF480372
VERSION CF480372.1
KEYWORDS GI:34509241
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 584)
AUTHORS Cordonnier-Pratt, M.-M., Suzukki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J., Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J. and Pratt, L.H.
TITLE EST database from Sorghum: pollen
JOURNAL Unpublished (2003)
COMMENT Other ESTs: POL1_65_H07_g1_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude

CF480372 584 bp mRNA linear EST 08-SEP-2003
POL1_65_H07_b1_A002 Pollen Sorghum bicolor cDNA clone
POL1_65_H07_A002 3', mRNA sequence.
CF480372
CF480372.1 GI:34509241
EST.
Sorghum bicolor (sorghum)
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 584)
Cordonnier-Pratt, M.-M., Suzukki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Eastman, A., Cannon, R., Kern, B., Morgan, J., Lucas, A., Al-Sheikh, A., Jones, V., Adibi, N., Owen, A., Gao, J. and Pratt, L.H.
EST database from Sorghum: pollen
Unpublished (2003)
Other ESTs: POL1_65_H07_g1_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude

vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.
Seq primer: Sug3-14 (TAGTCTAGCGCGCGGACG)
POLYA=yes.

FEATURES

source

Location/Qualifiers

1..584
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="POL1_65_H07_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Pollen"
/note="Organ: Pollen; Vector: pME18S-PL3; Site_1: XhoI; Site_2: XhoI; The library was prepared from polyA+ RNA from pollen at the late vacuolated-vacuolated stage of development. Pollen was harvested from greenhouse-grown panicles of sorghum line BTx623. Panicles were removed from the flag leaf prior to emergence, when no detectable amylase is present in pollen of male-fertile lines. This stage represents pollen collected from anthers about 8-14 days prior to anthesis. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-PL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Alignment Scores:

Pred. No.: 10.5 Length: 584
Score: 81.00 Matches: 24
Percent Similarity: 45.35% Conservative: 15
Best Local Similarity: 27.91% Mismatches: 37
Query Match: 17.23% Indels: 10
DB: 7 Gaps: 4

US-09-917-376-4 (1-89) x CF480372 (1-584)

QY 5 VallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
Db 45 ATAGCAATCAGCGGAACTCAAAATCAATCCGTCGGGAGGAGCGCTGTGTCGGCCGGC 104
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValArg--- 43
Db 105 -----CGCGCGGGAGACAGACCTGGAATCTTCAACGATCGGTGCAAGGAAA 152
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 153 CCATGGTGGGGATCGCAGGAGAGGTTTGACACCATGGAATATATATATATATGATGAA 212
QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
Db 213 GCTTGTGTGAAGTGTGGTGACGTGTTCTTTGGAGACGGTGTCTTGTCTGTGTCGCG 272
QY 79 AlaThrProThrAlaAsp 84
Db 273 AGCTTGAGTCTGGTGTAT 290

RESULT 2

BQ047524

LOCUS

DEFINITION EST596642 P. infestans-challenged potato leaf, incompatible reaction Solanum tuberosum cDNA clone BPL117P18 5' end, mRNA sequence.

ACCESSION

BQ047524

VERSION

BQ047524.1

KEYWORDS

EST.

SOURCE

Solanum tuberosum (potato)

ORGANISM

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiales; Solanales; Solanaceae; Solanum.

REFERENCE
AUTHORS

1 (bases 1 to 754)
Zhang, P., Hernandez, M., Tornqvist, C.-E., Wirtz, U., Loukoianov, A., Rangel, P., Haberlach, G. T., Karamycheva, S. A., Tsai, J., Chiemiango, A., Bougri, O., Buell, C. R., Ronning, C. M., Helgeson, J., and Baker, B.
Generation of ESTs from Potato Leaves Challenged with Phytophthora infestans, incompatible interaction (2002)
Unpublished (2002)

JOURNAL

COMMENT

Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T3.

FEATURES

source

Location/Qualifiers

1..754
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="BPL117P18"
/tissue_type="leaf"
/dev_stage="6 week old"
/lab_host="SOLR"
/clone_lib="P. infestans-challenged potato leaf, incompatible reaction"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: UC Berkeley, PGSC; sequencing: The Institute for Genomic Research. Whole plants were challenged with 450,000 sporangia/ml P. infestans isolate US-1 (US940501) in Biotron (Madison, Wisconsin). Leaf tissue was collected at 1, 2, 5, 12, and 24 hours post-challenge and frozen in liquid nitrogen immediately upon removal. Kennebec plants showed no signs of HR. Katahdin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."

ORIGIN

Alignment Scores:

Pred. No.: 22.5 Length: 754
Score: 79.50 Matches: 29
Percent Similarity: 38.05% Conservative: 14
Best Local Similarity: 25.68% Mismatches: 33
Query Match: 16.91% Indels: 37
DB: 5 Gaps: 4

US-09-917-376-4 (1-89) x BQ047524 (1-754)

QY 7 ValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGln 26
Db 152 GTGCTTATCAAGATCCAGGAAGT---CGGGGTAGTAATCTAATATCTCTCTGTTGAGTA 208
QY 27 LeuValAsnThrGlySerSer----- 34
Db 209 GTATCAAAATTCGGCACCTCTCACCTTTCCCTGGAAAATGCCCTATATATGAGTTTCGT 268
QY 35 -----ValAspLeuSerThrValThrValArgTyrTrpPheThr 47
Db 269 AAGGGGAGCCTCCGNAAGTTTCTTGGTTATGAACATTTCTCCACTCGCAATGGGTTC 328
QY 48 ArgAspGly-----GlySerSerThr 54
Db 329 AGGGTTGGCTCAGGATCATTCACACCAAGTGGCTGGGCTCTAGGCTTAGTCTTCTGAACT 388
QY 55 LeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGly 74
Db 389 CTGACCCCAAAATGGTGGGATTTCAAGGCTAGGTTCGTGTACTGTGACTCCAAATGGTGG 448
QY 75 SerValAsnProAlaThrProThrAlaAspThrTyrLeu 87


```

Db      449 GAA-----CCTCCTCCGAGATAGTTACCTT 475

RESULT 3
LOCUS   CB293085
DEFINITION
UCRCS01_05cb12_g1 Washington Navel orange cold acclimated flavedo &
albedo cDNA library Citrus sinensis cDNA clone UCRCS01_05cb12, mRNA
sequence.
CB293085
CB293085.1 GI:28618542
EST.
CITRUS sinensis
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
1 (bases 1 to 768)
Close,T.J., Collura,K., Fenton,R.D., Feuerbacher,O., Kim,H.R.,
Kudrna,D., Wanmaker,S., Wing,R. and Yu,Y.
Development of EST Resources and New Genetic Markers for California
Citrus
Unpublished (2003)
Contact: Timothy Close
Department of Botany & Plant Sciences, University of California
Riverside, CA, 92521-0124
Tel: 9097873318
Fax: 9097874437
Email: timothy.close@ucr.edu
Seq primer: T3.

FEATURES             Location/Qualifiers
     source
     1..768
         /organism="Citrus sinensis"
         /mol_type="mRNA"
         /cultivar="Washington navel"
         /db_xref="taxon:2711"
         /clone="UCRCS01_05cb12"
         /tissue_type="rind containing flavedo and albedo"
         /dev_stage="Mature fruit"
         /lab_host="E. coli TUC121"
         /clone_lib="Washington Navel orange cold acclimated
         flavedo & albedo cDNA library"
         /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
         Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
         field at University of California, Riverside Agricultural
         Operations since 1983. The scion was Washington Navel
         orange and the rootstock Carizzo Citrange. Tissue from
         mature fruit was harvested at mid-day in January 2002
         during a cold spell, when pre-dawn temperatures were
         approximately -2 to 2 degree C. Approximately 2 cm median
         sections of the rind were excised in the field from
         several fruits, then wrapped in aluminum foil and frozen
         quickly in dry ice. Total RNA was extracted using a phenol
         extraction procedure described in J. Japanese Soc. Hort.
         Sci. (1996) 64:809-814. Poly(A) RNA was purified, a cDNA
         library was made, and 1 million primary lambda cDNA clones
         were in vivo excised to give a population of pBluescript
         SK(-) phagemids. All steps to this point were performed in
         the TJ Close lab at the University of California,
         Riverside (Fenton). Phagemids were plated, plasmid DNA
         purified, cDNA clones archived, and DNA sequences
         determined bi-directionally using an ABI350 at the
         Arizona Genomics Institute, University of Arizona
         (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu).
         Chromatogram files were transmitted to UC Riverside (by
         Yu), then processed at UC Riverside (by Wanmaker) using
         the HarVest pipeline (http://harvest.ucr.edu) to remove
         vector and cloning oligo sequences and various
         contaminants, and to trim to a high quality region.
         Sequences that retained a phred 17 region of at least 100
         bases were deposited to GenBank."

ORIGIN
Alignment Scores:
Pred. No.:
58.7
Length:
906

US-09-917-376-4 (1-89) x CB293085 (1-768)

QY      3 GlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLys 22
      |||||:::
Db      612 GCGCGCATCAAC-----AAGAACACACAGCAGCAGCAATATATAAAGTAAT 662
      |||||:::
QY      23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42
      ::|||:::
Db      663 AATAATAGCAGCGTTGTACTAGTGGTAGCGTTGGTGGTCAAGTGGCGTGACG--- 719
      ::|||:::
QY      43 ArgTyrTrpPheThrArgAspGlyGlySerSerThr 54
      |||||:::
Db      720 -----TTTTCGAGAGCGGCGGAGGAGCAGT 746
      |||||:::

RESULT 4
LOCUS   BQ722991
DEFINITION
AGENCOURT 8103733 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:619I052 5', mRNA sequence.
BQ722991
BQ722991.1 GI:21861888
EST.
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 906)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA library Preparation: Life Technologies, Inc.
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13590 row: n column: 13
High quality sequence stop: 512.
Location/Qualifiers
     source
     1..906
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="IMAGE:619I052"
         /sex="male"
         /tissue_type="sympathetic trunk"
         /dev_stage="adult, 16 yr"
         /lab_host="DH10B"
         /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:
         NotI; Site 2: SalI; cDNA made by oligo-dT priming.
         Directionally cloned using the following adaptors:
         5'-TCGACCCACGCGCCG-3' and
         5'-GACTAGTCTTAGTCGCGACGCGCCCTT(15)-3'. Size selected >
         1 kb for average insert length 1.9 kb. This is a primary
         library, non-amplified. Library constructed by Life
         Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
         College of Medicine); available through Life
         Technologies."

ORIGIN
Alignment Scores:
Pred. No.:
58.7
Length:
906

```


Score: 77.00 Matches: 16
Percent Similarity: 49.15% Conservative: 13
Best Local Similarity: 27.12% Mismatches: 26
Query Match: 16.38% Indels: 4
DB: 5 Gaps: 2

US-09-917-376-4 (1-89) x BQ722291 (1-906)

Qy 7 ValGlnTyrLysAsnAspSerAlaProGlyAspGlnLleLysProGlyLeuGln 26
Db 548 CTGGCTATACCAATAAAATCTCCAGCGAGGCCACCGTGAGTA-----ATGAAG 598

Qy 27 LeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrpPhe 46
Db 599 GCTGTAAAGTGGGAATGAAGATATAGTTGGAAGCCTCTCCAGCACTACTGGTAC 658

Qy 47 ThrArgAspGlyLysSerThrLeuValTyrAsnCysAspTyrAlaAlaMetGly 65
Db 659 TCCGGG---GGCGGCATCGCGCCAGCTCTACACCTGCATCTGGGCGAGTGGTGA 712

RESULT 5
AA043716 376 bp mRNA linear EST 11-MAY-1997
LOCUS zK61f10.61 Soares pregnant uterus NBHPU Homo sapiens cDNA clone
DEFINITION IMAGE:487339 3' similar to gb:L10240 BASIGIN PRECURSOR (HUMAN),
mRNA sequence.

AA043716.1 GI:1521725
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 376)
Hallier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevasaki,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.

Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)

97044478
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 1400 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham

High quality sequence stop: 212.
Location/Qualifiers
1. .376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="CDB:3761053"
/db_xref="taxon:9606"
/clone="IMAGE:487339"
/sex="female"
/dev stage="adult"
/lab host="DH10B"

FEATURES
source
1. .376
/clone_lib="Soares_pregnant_uterus_NBHPU"
/note="Organ: uterus; Vector: pT73-Pac; Site1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AATGGAGAATTCGGCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I

ORIGIN

Alignment Scores:

Pred. No.: 25.2 Length: 376
Score: 76.00 Matches: 20
Percent Similarity: 43.42% Conservative: 13
Best Local Similarity: 26.32% Mismatches: 23
Query Match: 16.17% Indels: 20
DB: 1 Gaps: 3

US-09-917-376-4 (1-89) x AA043716 (1-376)

Qy 17 GlyAspAsnGlnLleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAsp 36
Db 64 GGGGGTTCACAGTGTAGCCCTGGT---GCTAGANAGCCGGCCGCTCTCCAGGGGTGAG 120

Qy 37 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyLysSerThr----- 54
Db 121 CTGGGGAGGCTT---CTGGGTTCTGGAGTCCCGCGGATGGCGCCAGTCCCGCAAAAC 177

Qy 55 -----LeuValTyrAsn 58
Db 178 CCCCCTCCAGAGTGTCCCGGATGCACAGAGAGGGGGCTTGGAGTGTACTTGAGGC 237

Qy 59 CysAspTyrAlaAlaMetGlyCysGlyAsnLleArgAlaSerPheGly 74
Db 238 TGTGACGGGTGCGCCCTCGGGTGTGGCAAGTCTCTCTGTGGGC 285

RESULT 6

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

US-09-917-376-4 (1-89) x AA043716 (1-376)

Qy 17 GlyAspAsnGlnLleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAsp 36
Db 64 GGGGGTTCACAGTGTAGCCCTGGT---GCTAGANAGCCGGCCGCTCTCCAGGGGTGAG 120

Qy 37 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyLysSerThr----- 54
Db 121 CTGGGGAGGCTT---CTGGGTTCTGGAGTCCCGCGGATGGCGCCAGTCCCGCAAAAC 177

Qy 55 -----LeuValTyrAsn 58
Db 178 CCCCCTCCAGAGTGTCCCGGATGCACAGAGAGGGGGCTTGGAGTGTACTTGAGGC 237

Qy 59 CysAspTyrAlaAlaMetGlyCysGlyAsnLleArgAlaSerPheGly 74
Db 238 TGTGACGGGTGCGCCCTCGGGTGTGGCAAGTCTCTCTGTGGGC 285

RESULT 6
LOCUS BE362891 606 bp mRNA linear EST 17-DEC-2001

DEFINITION DGI_90_C11.g1_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BE362891
VERSION BE362891.2 GI:15724434
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 606)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and
Pratt,L.H.

An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
On Jul 20, 2000 this sequence version replaced gi:9304448.

Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20. Three-prime sequences, which are obtained with PolyTMix or
T7 sequencing primer, are presented as the reverse complement.

Seq primer: PolyTMix
High quality sequence start: 15
High quality sequence stop: 594
POLYA=No.

Location/Qualifiers
1. 606
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site1: XhoI; Site2: EcoRI; The library was

FEATURES
source
1. 606
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site1: XhoI; Site2: EcoRI; The library was

made from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."

ORIGIN

Alignment Scores:
Pred. No.: 46.7 Length: 606
Score: 76.00 Matches: 23
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match: 16.17% Indels: 10
DB: 2 Gaps: 4

US-09-917-376-4 (1-89) x BE362891 (1-606)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
Db 208 ATAGCAATCAGCGGAATCAAAATCGGTCCGGAGGAGCGTGTGGCCCGGC 267
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
Db 268 -----CGCGCCGGGACAGAACCTTCAAACTTCCAAACGATCGGTGCAAGGAAA 315
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerThrLeuValThrValAsnCysAspTrp 61
Db 316 CCATGTTGGGGATCTGCGGGAGGAGGTTTGAGCAACATGGAATATATATATGATGAA 375
QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
Db 376 GCTTGTGTGAAGTGTGGTGTGCTTTTGGAGAGCGTGTCTTGTCTGTGGTCCG 435
QY 79 AlaThrProThrAlaAsp 84
Db 436 AGCTTGCAGTCTGGTGAT 453

RESULT 7

CD226529 641 bp mRNA linear EST 21-MAY-2003
LOCUS CCC1_46_E07_b1_A007 Callus culture/cell suspension Sorghum bicolor
DEFINITION CDNA clone CCC1_46_E07_A007 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2003)
Contact: Cordonnier-Pratt MM
Other ESTs: CCC1_46_E07_g1_A007
An EST database from Sorghum: callus culture and cell suspension
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science, Plant Material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.
Seq primer: Sug3 (CGACCTGCAGCTCGACACA)

FEATURES

source

Location/Qualifiers

1. .641

/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="RTx430"
/db_xref="taxon:4558"
/clone="CCC1_46_E07_A007"
/lab_host="DH10B-r1-phage-resistant E. coli"
/clone_lib="Callus culture/cell suspension"
/note="Vector: pME18S-FL3; Site_1: XhoI; Site_2: XhoI; The
library was prepared from a mixture of polyA+ RNA from
callus culture tissue and cells in suspension culture.
Double-stranded cDNA was cloned unidirectionally into
different DraIII sites of the pME18S-FL3 vector (5-prime
DraIII site is CACTGTGTG, 3-prime DraIII site is
CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Alignment Scores:
Pred. No.: 50.2 Length: 641
Score: 76.00 Matches: 23
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match: 16.17% Indels: 10
DB: 6 Gaps: 4

US-09-917-376-4 (1-89) x CD226529 (1-641)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
Db 149 ATAGCAATCAGCGGAATCAAAATCGGTCCGGAGGAGCGTGTGGCCCGGC 208
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
Db 209 -----CGCGCCGGGACAGAACCTTCCAAACGATCGGTGCAAGGAAA 256
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerThrLeuValThrValThrValArg 61
Db 257 CCATGTTGGGGATCTGCGGGAGGAGGTTTGAGCAACATGGAATATATATGATGAA 316
QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
Db 317 GCTTGTGTGAAGTGTGGTGTGCTTTTGGAGAGCGTGTCTTGTCTGTGGTCCG 376

RESULT 8

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: PolyTMix
High quality sequence start: 89
High quality sequence stop: 651
POLYA=No.

FEATURES

source

Location/Qualifiers
1. .657
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DG1)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

ORIGIN

Alignment Scores:
Pred. No.: 51.8 Length: 657
Score: 76.00 Matches: 23
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match: 16.17% Indels: 10
DB: 2 Gaps: 4

US-09-917-376-4 (1-89) x BE362820 (1-657)

QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleIysProGly 24
::: :
DB 204 ATAGCAATCAGCCGAAATCAATGAATCCGGTCCGGAGGAGCGTGTGCGCCGGC 263
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg-- 43
DB 264 -----CGGCGCGGACAGACCCCTGAACTCCAAACGATCGTGCGCAGGAAA 311
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61
::: :
DB 312 CCATGGTGGGGATCGCAGGAGGAGGTTTGAGCAACATGAATATATATATGATGAA 371
QY 62 AlalaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
::: :
DB 372 GCTTGTGGAAGTGTGGTGTGGTCTTTTGGAGACGGTGTCTTGTCTGTGTGTCG 431
QY 79 AlaThrProThrAlaAsp 84
::: :
DB 432 AGCTTGAGCTCTGGTGAT 449

RESULT 9

CF487649

LOCUS

CF487649 676 bp mRNA linear EST 08-SEP-2003
POLI_45_C08.bl_A002 Pollen Sorghum bicolor cDNA clone

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Sorghum bicolor (sorghum)
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE

AUTHORS

1 (bases 1 to 676)
Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,
Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J.,
Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and
Pratt,L.H.

TITLE

JOURNAL

COMMENT

EST database from Sorghum: pollen
Unpublished (2003)
Other_ESTs: POLI_45_C08.g1_A002

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.

Seq primer: Sug3-14 (TAGTCTAGCGCGCGGAC)

POLYA=yes.

FEATURES

source

Location/Qualifiers

1. .676
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="POLI_45_C08_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Pollen"
/note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI;
Site 2: XhoI; The library was prepared from polyA+ RNA
from pollen at the late vacuolated-vacuolated stage of
development. Pollen was harvested from greenhouse-grown
panicles of sorghum line BTx623. Panicles were removed
from the flag leaf prior to emergence, when no detectable
amylase is present in pollen of male-fertile lines. This
stage represents pollen collected from anthers about 8-14
days prior to anthesis. Double-stranded cDNA was cloned
unidirectionally into different DraIII sites of the
pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG,
3-prime DraIII site is CACCATGTG). XhoI excises the cDNA
insert."

ORIGIN

Alignment Scores:
Pred. No.: 53.7 Length: 676
Score: 76.00 Matches: 23
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match: 16.17% Indels: 10
DB: 7 Gaps: 4

US-09-917-376-4 (1-89) x CF487649 (1-676)

QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleIysProGly 24
::: :
DB 214 ATAGCAATCAGCCGAAATCAATGAATCCGGTCCGGAGGAGCGTGTGCGCCGGC 273
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg-- 43
DB 274 -----CGGCGCGGACAGAACCTTCCAAACGATCGTGCGCAGGAAA 321
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61
::: :
DB 322 CCATGGTGGGGATCGCAGGAGGAGGTTTGAGCAACATGAATATATATATGATGAA 381
QY 62 AlalaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
::: :
DB 382 GCTTGTGGAAGTGTGGTGTGGTCTTTTGGAGACGGTGTCTTGTCTGTGTGTCG 441
QY 79 AlaThrProThrAlaAsp 84
::: :
DB 442 AGCTTGAGCTCTGGTGAT 459

RESULT 10
BG739392 691 bp mRNA linear EST 15-MAY-2001
LOCUS
DEFINITION
EM1_82_D05.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION
BG739392
BG739392.1 GI:14089081
VERSION

EST.
Sorghum bicolor (sorghum)
Sorghum bicolor
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 691)
Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: PolyTMix
High quality sequence start: 50
High quality sequence stop: 691
POLYA=No.

FEATURES

Location/Qualifiers
1..691

/organism="Sorghum bicolor"

/mol_type="mRNA"

/db_xref="taxon:4558"

/clone_lib="Embryo 1 (EM1)"

/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from lambda Zap II; Site 1: XhoI; Site 2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."

ORIGIN

Alignment Scores:
Pred. No.: 55.3 Length: 691
Score: 76.00 Matches: 23
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match: 16.17% Indels: 10
DB: 4 Gaps: 4

US-09-917-376-4 (1-89) x BG739392 (1-691)

QY 5 ValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
Db 265 ATAGCAATCAGCCGAAACTCAATGAATCCGGTCCGGAGGAGCCTGTGTCGCCCGGC 324
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
Db 325 -----CGCGCCGGGACAGAACCTGAACCTTCAACCATCGTGCAGAGAA 372
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 373 CCATGGTGGGGATCTGCAGGAGGAGGTTTGAGCAACATGGAATATATATATATGATGAA 432
QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
Db 433 GCTTGTGTGAAGTGTGGTGACGTGTGTTTGTGGAGCGGTGTCTGTGCTCTGTGTCG 492
QY 79 AlaThrProThrAlaAsp 84
Db 493 AGCTTGCAGTCTGTGTAT 510

RESULT 11

BZ616323 724 bp DNA linear GSS 16-JAN-2003
LOCUS BZ616323/c
DEFINITION 1958e08.g1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
1958e08 5', genomic survey sequence.
ACCESSION BZ616323

VERSION

BZ616323.1 GI:27766878

GSS.

KEYWORDS

Ze mays

SOURCE

Ze mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 724)

AUTHORS

Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N.,
Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zutavern, T., McCombie, W.R. and Martienssen, R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered)
Unpublished (2002)

TITLE

JOURNAL

COMMENT

Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mccombie@cshl.org
Plate: ig58 row: e column: 08
Seq primer: -21M13UnivRev
Class: shotgun
High quality sequence stop: 728.

FEATURES

Location/Qualifiers

1..724

/organism="Zea mays"

/mol_type="genomic DNA"

/cultivar="B73"

/db_xref="taxon:4577"

/clone="ig58e08"

/lab_host="DH5a"

/clone_lib="WGS-ZmaysF (DH5a methyl filtered)"

/note="Organ: Immature ears; Site 1: Xba I; Site 2: Xba I;
The vector was digested with XbaI and one nucleotide was
added by fill in the recessive 3' end. The genomic DNA
was nebulized, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector (.x/y
reads in M13mp19, .b/g reads in pUC19). The same ligation
was transformed into DH5a."

ORIGIN

Alignment Scores:
Pred. No.: 58.7 Length: 724
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.17% Indels: 43
DB: 6 Gaps: 6

US-09-917-376-4 (1-89) x BZ616323 (1-724)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAsnAspSer-----AlaPro 16
Db 587 TCGCAGGGGTGAGCATCCATACAGAGGCGGCGACGACGCTGGCGAGTGGCGCC 528
QY 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db 527 GCGCGCCAGCTGAGCAGCTGCAGCGAGGACGACGATGCTTATCAACCGTGGCGC 468
QY 30 Thr-GlySerSerSerValAsp----- 36
Db 467 ACCGAGGAAGTTCGTCAACTTCCGAGATCGTGGAGCGGTCCAGAGGCTGGGTTCGAG 408
QY 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlySerSe 53
Db 407 GTCATCCGATCGAGCGGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 348
QY 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db 347 TCGTGCG-----ACGTGCTCATGGCGCGCACGGG----- 317


```

Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 316 -----CCGGCCTCACCAACTTCTCTTCC 293

RESULT 12
BI218457
LOCUS 602937951F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5101069 5',
DEFINITION mRNA sequence.
ACCESSION BI218457
VERSION BI218457.1 GI:14671901
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1. (bases 1 to 814)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1A11243 row: f column: 14
High quality sequence stop: 792.
FEATURES
Location/Qualifiers
1..814
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5101069"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Li9"
/notes="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
ORIGIN
Alignment Scores:
Pred. No.: 68.3 Length: 814
Score: 76.00 Matches: 28
Percent Similarity: 44.83% Conservatve: 11
Best Local Similarity: 32.18% Mismatches: 29
Query Match: 16.17% Indels: 19
DB: 4 Gaps: 4
US-09-917-376-4 (1-89) x BI218457 (1-814)
Qy 2 SerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnle 21
Db 33 GCAGGTGGAGCCCGAGTATGGGAGCGGAACACACAGATCGCCAGGACAATGAGCAG--- 89
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 90 -----AGTTCCCGAGTGGACCTGAGGACCCTG--- 116
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeu-----ValTyrAsnCys 59
Db 117 CTCGCTACTACACACAGAGCGCGCGCTCTCACAGATCCAGCGAGTACGGCTGT 176
Qy 60 Asp-TrpAlaAlaMetGlyCysGlyAsnleArgAlaSerPheGlySerValAsnProAl 79
Db 177 GACGTGGGTTCGAGCGGCGCC-----TCCTCCGCGGTACGAGCAGGTTCGCA 224
Qy 79 aThrProThrAlaAspThr 85

```

```

Db 225 TACGACGGCTCGGATTACA 243
RESULT 13
BZ545807
LOCUS OGALR72TM ZM2_0.7_1.5_KB Zea mays genomic clone ZMMBMA0089L24,
DEFINITION genomic survey sequence.
ACCESSION BZ545807
VERSION BZ545807.1 GI:27099699
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 905)
REFERENCE
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Hedell,J.A., Rohlfing,T.,
Citek,R.W., Numborg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other GSSs: OGALR72TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
1..905
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0089L24"
/clone_lib="ZM2_0.7_1.5_KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN
Alignment Scores:
Pred. No.: 78.3 Length: 905
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservatve: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.17% Indels: 43
DB: 8 Gaps: 6
US-09-917-376-4 (1-89) x BZ545807 (1-905)
Qy 2 SerGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 87 TCCGACGGCTGAGCATCCATCCATACAGGAGGCGGACAGCAGCTGCGGAGCTGGCCCC 146
Qy 17 Gly-----AspAsnGlnleLysProGlyLeuGlnLeuValAsn----- 29
Db 147 GCGCGCCCGAGCTGAGCAGCTGCGCGGAGGAGCCAGCAGTGTATCAACCGTGGCCCG 206
Qy 30 Thr-GlySerSerValAsp----- 36
Db 207 ACAGAGGAGTTCGTCACTCCCGAGATGTCGAGGCGGTCCAGAGGCTGGGTTCGAG 266
Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db 267 GTCATCCCGATCGAGCGCGGTAGACCTCAGCGTGGAGGACTTGGCCGCGACCGTGCAC 326
Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnleArgAlaSerPh 73
Db 327 TCGTGGC-----ACGTGCTCATGGCGCGCAGCGGG----- 357

```



```

Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 358 -----CGGGCTCACCACACTTCTTCTTCC 381

RESULT 14
CG270741/c
LOCUS CG270741 934 bp DNA linear GSS 25-AUG-2003
DEFINITION OGWIB33TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0583E18,
genomic survey sequence.
ACCESSION CG270741
VERSION CG270741.1 GI:34182882
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 934)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
JOURNAL Other GSSs: OGWIB33TV
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..934
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0583E18"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 81.5 Length: 934
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.17% Indels: 43
DB: 9 Gaps: 6

US-09-917-376-4 (1-89) x CG270741 (1-934)
Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 804 TCGCGAGGGCTCAGCATCCATACAGGAGGGCAACAGCAACGCTCGCGAGCTGGCGCC 745
Qy 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db 744 GGGCCCCAGCTGAGCAGCTGAGCGAGGAGCAACGACGACTGATGCTTATCAACCGTGGCCGC 685
Qy 30 Thr-GlySerSerValAsp----- 36
Db 684 ACAGGAGAGTTGCTCAACTCCCGAGATCGTCGGAGCGGTCCAGAGGCTGGGTTCCGAG 625
Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db 624 GTCATCCCGATCGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCCGAC 565
Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db 564 TCGTGCG-----ACGTGCTCATGGCGCGCACGGGG----- 534

```

```

Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 533 -----CGGGCTCACCACACTTCTTCTTCC 510

RESULT 15
CG270748
LOCUS CG270748 1001 bp DNA linear GSS 25-AUG-2003
DEFINITION OGWIB33TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0583E18,
genomic survey sequence.
ACCESSION CG270748
VERSION CG270748.1 GI:34182889
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
JOURNAL Other GSSs: OGWIB33TH
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..1001
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0583E18"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 89.1 Length: 1001
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.17% Indels: 43
DB: 9 Gaps: 6

US-09-917-376-4 (1-89) x CG270748 (1-1001)
Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 418 TCGCGAGGGCTCAGCATCCATACAGGAGGGCAACAGCAACGCTCGCGAGCTGGCGCC 477
Qy 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db 478 GGGCCCCAGCTGAGCAGCTGAGCGAGGAGCAACGACGACTGATGCTTATCAACCGTGGCCGC 537
Qy 30 Thr-GlySerSerValAsp----- 36
Db 538 ACAGGAGAGTTGCTCAACTCCCGAGATCGTCGGAGCGGTCCAGAGGCTGGGTTCCGAG 597
Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db 598 GTCATCCCGATCGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCCGAC 657
Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73

```


Db 658 TCGTGG-----ACGTGCTCATGGGCGGCACGGGG----- 688
Qy 73 eGlySerValAsnProAlaThrProThrAlaaspThr 85
Db 689 -----CCGGGCTCACCAACTTCTTCTCC 712

Search completed: October 5, 2005, 21:09:32
Job time : 860.806 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 07:47:30 ; Search time 42.3629 Seconds
(without alignments)
3437.649 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQVQKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US0917376/runat 04102005.164328 26925/app query.fasta 1.2524
-DB=Issued Patents NA -QWTF=fastap -SUPFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0917376 @CGN 1 1 516 @runat 04102005.164328 26925 -NCFU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.6	2029	3 US-09-136-574A-46	Sequence 46, Appl
2	194.5	41.4	6416	3 US-09-136-574A-2	Sequence 2, Appli
3	192.5	41.0	11707	3 US-09-136-574A-1	Sequence 1, Appli
4	182.5	38.8	2977	2 US-07-862-588B-1	Sequence 1, Appli
5	165	35.1	1438	4 US-09-339-159B-3	Sequence 3, Appli
6	165	35.1	1482	3 US-09-198-956-9	Sequence 9, Appli
7	165	35.1	1482	3 US-09-198-955A-11	Sequence 11, Appl
8	165	35.1	1482	3 US-09-694-531-11	Sequence 11, Appl
9	165	35.1	1482	3 US-09-670-141-9	Sequence 9, Appli
10	165	35.1	1482	4 US-10-072-152-11	Sequence 11, Appl
11	155.5	33.1	4059	4 US-09-784-554B-1	Sequence 1, Appli
12	150.5	32.0	4056	4 US-09-784-554B-3	Sequence 3, Appli

13	120.5	25.6	1624	2 US-07-862-588B-6	Sequence 6, Appli
14	120.5	25.6	1775	1 US-07-862-588B-5	Sequence 5, Appli
15	114	24.3	486	1 US-08-048-164A-1	Sequence 1, Appli
c 16	114	24.3	486	1 US-08-048-164A-3	Sequence 3, Appli
c 17	114	24.3	486	1 US-08-460-462-1	Sequence 1, Appli
c 18	114	24.3	486	1 US-08-460-462-3	Sequence 3, Appli
c 19	114	24.3	486	1 US-08-460-457-1	Sequence 1, Appli
c 20	114	24.3	486	1 US-08-460-457-3	Sequence 3, Appli
c 21	114	24.3	486	1 US-08-460-458-1	Sequence 1, Appli
c 22	114	24.3	486	2 US-08-460-455-1	Sequence 1, Appli
c 23	114	24.3	486	2 US-08-460-455-3	Sequence 3, Appli
c 24	114	24.3	486	2 US-08-330-394A-1	Sequence 1, Appli
c 25	114	24.3	486	2 US-08-330-394A-3	Sequence 3, Appli
c 26	114	24.3	486	3 US-09-006-636-5	Sequence 5, Appli
c 27	114	24.3	499	3 US-09-006-632-5	Sequence 5, Appli
c 28	114	24.3	499	3 US-09-325-274-5	Sequence 5, Appli
c 29	114	24.3	499	3 US-09-277-716-21	Sequence 21, Appl
c 30	113	24.0	1146	3 US-09-609-161B-21	Sequence 21, Appl
c 31	113	24.0	1146	3 US-09-335-586-3	Sequence 3, Appli
c 32	77	16.4	8107	4 US-09-949-016-12565	Sequence 12565, A
c 33	76	16.2	136058	4 US-09-949-016-17064	Sequence 17064, A
c 34	76	16.1	5714	4 US-09-949-016-17064	Sequence 17064, A
c 35	75.5	16.1	5714	4 US-09-949-016-17064	Sequence 17064, A
c 36	72.5	15.4	819	4 US-09-902-540-3752	Sequence 3752, Ap
c 37	72.5	15.4	17897	4 US-09-902-540-1182	Sequence 1182, Ap
c 38	72.5	15.4	20284	4 US-09-526-193A-21	Sequence 21, Appl
c 39	70	14.9	10892	4 US-09-902-540-962	Sequence 962, App
c 40	70	14.9	25165	3 US-09-453-702B-39	Sequence 39, Appl
c 41	70	14.9	246444	4 US-09-949-016-13113	Sequence 13113, A
c 42	69.5	14.8	103987	4 US-09-949-016-12513	Sequence 12513, A
c 43	69.5	14.8	103988	4 US-09-949-016-17050	Sequence 17050, A
c 44	68.5	14.6	601	4 US-09-949-016-128395	Sequence 128395,
c 45	68.5	14.6	601	4 US-09-949-016-128395	Sequence 128395,

ALIGNMENTS

RESULT 1
US-09-136-574A-46
; Sequence 46, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSER: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136.574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

Thu Oct 6 11:20:46 2005

us-09-917-376-4.p2n.rni

STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11707 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
IS-09-136-574A-1

4

Db 1051 ATCAATCCTCAGTTCAAGGTTACTAATACCGAAGCAGTGAATGATTTCTCCAACTC 1110
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGAGCAGAAAGATCAGACCTTCTGG---TGTGAC 1167
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGGCAGCTACAACGGAATTTACTTCAAAATGTAAGA 1227
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGAAAAATGAGTTCTCTCAACAAATAACGAGCAGACCTACCTTTGAA 1278

RESULT 10

US-10-072-152-11

; Sequence 11, Application US/10072152
; Patent No. 6677147

; GENERAL INFORMATION:

; APPLICANT: Andersen, Lene N.

; APPLICANT: Schulein, Martin

; APPLICANT: Lange, Niels E.

; APPLICANT: Bjornvad, Mads E.

; APPLICANT: Moller, Soren

; APPLICANT: Glad, Samne O. S.

; APPLICANT: Kauppinen, Markus S.

; APPLICANT: Schnorr, Kirk

; APPLICANT: Kongsbak, Lars

; TITLE OF INVENTION: No. 6677147el Pectate Lyases

; FILE REFERENCE: 5378.200-US

; CURRENT APPLICATION NUMBER: US/10/072.152

; CURRENT FILING DATE: 2002-02-07

; PRIOR APPLICATION NUMBER: US/09/198,955

; PRIOR FILING DATE: 1998-11-24

; PRIOR APPLICATION NUMBER: 1343/97

; PRIOR FILING DATE: 1997-11-24

; PRIOR APPLICATION NUMBER: 1344/97

; PRIOR FILING DATE: 1997-11-24

; PRIOR APPLICATION NUMBER: 60/067,249

; PRIOR FILING DATE: 1997-12-02

; PRIOR APPLICATION NUMBER: 60/067,240

; PRIOR FILING DATE: 1997-12-02

; PRIOR APPLICATION NUMBER: 09/073,684

; PRIOR FILING DATE: 1998-05-06

; PRIOR APPLICATION NUMBER: 09/184,217

; PRIOR FILING DATE: 1998-11-02

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 1482

; TYPE: DNA

; ORGANISM: Clostridium thermocellum

US-10-072-152-11

Alignment Scores:

Pred. No.:	6,97e-13	Length:	1482
Score:	165.00	Matches:	35
Percent Similarity:	59.79%	Conservative:	23
Best Local Similarity:	36.08%	Mismatches:	29
Query Match:	35.11%	Indels:	10
DB:	4	Gaps:	2

US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)

Qy 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTGGAAGTTGTAATCTCAACAGCAATCTTCAGATACTACTAATCA 1050
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGGTTACTAATAACCGAAGCAGTGAATGATTTCTCCAACTC 1110
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

Db 1111 ACATTGAGATATTATTATACAGTACGAGCAGAAAGATCAGACCTTCTGG---TGTGAC 1167
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGGCAGCTACAACGGAATTTACTTCAAAATGTAAGA 1227
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGAAAAATGAGTTCTCTCAACAAATAACGAGCAGACCTACCTTTGAA 1278

RESULT 11

US-09-784-554B-1

; Sequence 1, Application US/09784554B

; Patent No. 6815192

; GENERAL INFORMATION:

; APPLICANT: Schnorr, Kirk

; APPLICANT: Jorgensen, Per Lina

; APPLICANT: Schulein, Martin

; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES

; FILE REFERENCE: 10017.200-US

; CURRENT APPLICATION NUMBER: US/09/784,554B

; CURRENT FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 4059

; TYPE: DNA

; ORGANISM: Paenibacillus polymyxa

US-09-784-554B-1

Alignment Scores:

Pred. No.:	6,2e-11	Length:	4059
Score:	155.50	Matches:	27
Percent Similarity:	56.47%	Conservative:	21
Best Local Similarity:	31.76%	Mismatches:	36
Query Match:	33.09%	Indels:	1
DB:	4	Gaps:	1

US-09-917-376-4 (1-89) x US-09-784-554B-1 (1-4059)

Qy 4 GlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3598 GGACTGCTGCTCCAAATATCGCACAGCAGATATAAGGTGAACGATAATCACCTCAATCCG 3657
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArg 43
Db 3658 CAATTCCAAATTTGTAACAACAAAGGCACCACTCCATACCGATCAACGAGTTGAAAATTCGC 3717
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 3718 TACTACTACAACTCGACGGTGACCGTGAGCAG---ACTTTCACTGCCACTATGCGACG 3774
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3775 CTGAGCTGCTCAAGCTGAACGGTAACTGGTTAAATCGAAGGCTCAACCGGTGCC 3834
Qy 84 AspThrTyrLeuGln 88
Db 3835 GATTATTATTGGAA 3849

RESULT 12

US-09-784-554B-3

; Sequence 3, Application US/09784554B

; Patent No. 6815192

; GENERAL INFORMATION:

; APPLICANT: Schnorr, Kirk

; APPLICANT: Jorgensen, Per Lina

; APPLICANT: Schulein, Martin

; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES

; FILE REFERENCE: 10017.200-US

; CURRENT APPLICATION NUMBER: US/09/784,554B

; CURRENT FILING DATE: 2001-02-16


```
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 3.06e-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.4% Conservative: 21
Best Local Similarity: 31.76% Mismatches: 36
Query Match: 32.02% Indels: 1
DB: 4 Gaps: 1

US-09-917-376-4 (1-89) x US-09-784-554B-3 (1-4056)
QY 4 GlyVallyValGlnTyrIleAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 3595 GGATTGGTCTCCAGTATCGCAGCGGATCAATGTGAAGCAATCACTTGAACCG 3654
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 3655 CATTTCCTCAATTTAAATAAGGTACATCTCGTACCGATCAACGAGTGTGAAAATTGCG 3714
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnGlyAspTrpAlaAla 63
DB 3715 TACTACTACACGATCGCGTGCAGCGTGACGAGCAG---ACATTCACTGCGACTATCGGTG 3771
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 3772 CTGACTCTCTCGAAGCTGAATGTTAGTGGTTAAATGGATAAAGCTGCAACCGGTGCT 3831
QY 84 AspThrThrLeuGln 88
DB 3832 GATTATTATTGGAA 3846

RESULT 13
US-07-862-588B-6
; Sequence 6, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5916796 of No. 5916796th America, Inc.
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728

; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1607
; OTHER INFORMATION:
US-07-862-588B-6

Alignment Scores:
Pred. No.: 1.18e-06 Length: 1624
Score: 120.50 Matches: 28
Percent Similarity: 54.12% Conservative: 18
Best Local Similarity: 32.94% Mismatches: 36
Query Match: 25.64% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-4 (1-89) x US-07-862-588B-6 (1-1624)
QY 2 SerGlyGlyVallyValGlnTyrIleAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 1323 AGGGGAACCTTGTTCATACAAAGTGGCGACACTAGCCGCGGATACCAATG 1382
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 1383 AGCCCTTCCTTTAATCAATCAAGACACGCTGACACCCCTGTTAACTGAGCGGCTCAAG 1442
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnGlyAspTrp 61
DB 1443 CTTNNNNNNNNNNNNNNAAAAAGAC---GGACCTGCGGATATGAGCTGCTCGATCGACTGG 1499
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 1500 GCGCAATCGCGCAACGAATGTTCTGCTGGCATTC-----GCTAACTTACGGGAGT 1553
QY 82 ThrAlaAspThrTyr 86
DB 1554 AATACGATACCTTAC 1568

RESULT 14
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5916796 of No. 5916796th America, Inc.
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
```


;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/862,588B

;; FILING DATE: 19920727

;; CLASSIFICATION: 435

;; PRIORITY APPLICATION DATA:

;; APPLICATION NUMBER: DK 164/90

;; FILING DATE: 19-JAN-1990

;; PRIORITY APPLICATION DATA:

;; APPLICATION NUMBER: PCT/DK91/00013

;; FILING DATE: 18-JAN-1991

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Zelson, Steve T. / Lambiris, Elias J.

;; REGISTRATION NUMBER: 30,335 / 33,728

;; REFERENCE/DOCKET NUMBER: 3425.204-US

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 212 867 0123

;; TELEFAX: 212 867 0298

;; INFORMATION FOR SEQ ID NO: 5:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 1775 base pairs

;; TYPE: NUCLEIC ACID

;; STRANDEDNESS: double

;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)

;; FRAGMENT TYPE: N-terminal

;; ORIGINAL SOURCE:

;; ORGANISM: Bacillus lautus

;; STRAIN: NCIMB 40250

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: 30..(1625,1775)

;; IDENTIFICATION METHOD: experimental

;; OTHER INFORMATION: /partial

;; OTHER INFORMATION: /evidence= EXPERIMENTAL

;; OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:

US-07-862-588B-5

Alignment Scores:
Pred. No.: 1.34e-06 Length: 1775
Score: 120.50 Matches: 28
Percent Similarity: 54.12% Conservative: 18
Best Local Similarity: 32.94% Mismatches: 36
Query Match: 25.64% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-4 (1-89) x US-07-862-588B-5 (1-1775)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 1323 ACGGGAAACCTGTGTTCCATACAAAGTTGGGACACTACGGCCACGGATAACCAATG 1382
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 1383 AAGCTTCTCTTAACATCAAGAACCAACGGTACACCCCTGTTAACTTGAGGGCTCAAG 1442
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 1443 CTTNNNNNNNNNNNNAAAAAGAC---GGACCTGGGATATGAGTGTCTCGACTG 1499
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 1500 GCGAAATCGCGCGAAGCAATGTTCTGTGTCGATTC-----GCTAACTTTACCGGGAGT 1553
QY 82 ThrAlaAspThrTyr 86
DB 1554 AATACGGATACTTAC 1568

RESULT 15

US-08-048-164A-1

;; Sequence 1, Application US/08048164A

;; Patent No. 5496934

;; GENERAL INFORMATION:

;; APPLICANT: Shoseyov, Oded
;; APPLICANT: Shpiegl, Itai
;; APPLICANT: Goldstein, Marc A.
;; APPLICANT: Doi, Roy H.
;; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
;; NUMBER OF SEQUENCES: 21
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: PENNIE & EDMONDS
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/048,164A
;; FILING DATE: 14-APR-1993
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Mistrock, S. Leslie
;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 7809-003
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-8864/9741
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 486 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..486
US-08-048-164A-1

Alignment Scores:
Pred. No.: 1.65e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservative: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 24.26% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-4 (1-89) x US-08-048-164A-1 (1-486)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 7 ACATCATCAATGTCAGTTGAATTTTCAACTCTAAACAAATCAGCACAAACAACTCAATT 66
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 67 ACACCAATTAATCAAAATTACTAACACATCTGACAGTGATTTAAATTTAAATGACGTAAA 126
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 127 GTTAGATATTATTACACAAAGTGATGTACACAGGACAAACTTTCTGG---TGTGACCAT 183
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
DB 184 CTTGGTGCATTATTAGGAAATAGCTATTGTTGATAACACTAGCAAGTGACAGCAAACTTC 243
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
DB 244 GTTAAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 291

Search completed: October 5, 2005, 22:56:54
Job time : 49.3629 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 08:30:02 ; Search time 171.446 Seconds
(without alignments)
3584.301 Million cell updates/sec

Title: US-09-917-376-4
Perfect score: 470
Sequence: 1 VSGGVKQYKKNDSAPGDNQ.....RASFGSYNPATPTADTYLQX 89

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7442561 seqs, 3452328358 residues
Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09917376/runat_04102005_164329_26954/app.query.fasta_1.2524
-DB=Published Applications NA -QWTF=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-TRANSCU=0 -LOOPEXT=0 -UNIT5=bits -START=1 -ENR=-1 -MATRIX=blosum62
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USPR=US09917376 @CGN 1 1 2178 @runat 04102005_164329_26954
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LANGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	469	99.8	2289	10	US-09-917-378-2	Sequence 2, Appli
2	466	99.1	3687	10	US-09-917-384-2	Sequence 2, Appli
3	466	98.1	3687	10	US-09-917-383-2	Sequence 2, Appli
4	462	98.3	2869	11	US-09-917-376-2	Sequence 2, Appli
5	462	98.3	2869	15	US-10-155-400-2	Sequence 2, Appli
6	257	54.7	2223	15	US-10-156-761-550	Sequence 550, App
7	257	54.7	9025608	15	US-10-156-761-1	Sequence 11, Appli
8	190	40.4	1527	17	US-10-369-493-46838	Sequence 46838, A
9	188	40.0	1434	19	US-10-466-208-11	Sequence 11, Appli
10	188	40.0	1491	19	US-10-466-208-7	Sequence 7, Appli
11	188	40.0	2510	19	US-10-433-577-34	Sequence 34, Appli
12	165	35.1	501	21	US-10-933-404-1	Sequence 1, Appli
13	165	35.1	1438	17	US-10-372-054-3	Sequence 3, Appli
14	165	35.1	1482	13	US-10-072-152-11	Sequence 11, Appli
15	165	35.1	1482	18	US-10-655-433-11	Sequence 11, Appli
16	157	33.4	1314	9	US-09-974-300-684	Sequence 684, App
17	155.5	33.1	4059	10	US-09-784-554B-1	Sequence 1, Appli
18	155.5	33.1	4059	20	US-10-896-555-1	Sequence 1, Appli
19	150.5	32.0	4056	10	US-09-784-554B-3	Sequence 3, Appli
20	150.5	32.0	4056	20	US-10-896-555-3	Sequence 3, Appli
21	131.5	28.0	3180	22	US-10-886-393A-29	Sequence 29, Appli
22	130.5	27.8	4452	17	US-10-282-122A-15299	Sequence 15299, A
23	126.5	26.9	1303	21	US-10-490-737-1	Sequence 1, Appli
24	126.5	26.9	1747	21	US-10-490-737-2	Sequence 2, Appli
25	126.5	26.9	3489	21	US-10-490-737-3	Sequence 3, Appli
26	114	24.3	768	16	US-10-261-446-5	Sequence 5, Appli
27	114	24.3	768	18	US-10-261-445B-5	Sequence 5, Appli
28	113	24.0	768	21	US-10-782-234-5	Sequence 5, Appli
29	113	24.0	652	20	US-10-748-055-9	Sequence 9, Appli
30	113	24.0	1146	10	US-09-808-898-21	Sequence 21, Appli
31	78.5	16.7	764	20	US-10-653-047-5404	Sequence 5404, Ap
32	78.5	16.7	3147	17	US-10-282-122A-13716	Sequence 13716, A
33	76	16.2	757	19	US-10-767-701-8547	Sequence 8547, Ap
34	76	16.2	96597	18	US-10-052-482-226	Sequence 226, App
35	75.5	16.1	1041	17	US-10-260-238-742	Sequence 742, App
36	75.5	16.1	1041	19	US-10-437-963-76751	Sequence 76751, A
37	75.5	16.1	2091	18	US-10-211-462-222	Sequence 222, App
38	75.5	16.1	2091	20	US-10-723-860-4038	Sequence 4038, Ap
39	75.5	16.1	2091	22	US-10-756-149-4006	Sequence 4006, Ap
40	75.5	16.1	3396	17	US-10-398-458-1	Sequence 1, Appli
41	75.5	16.1	4041	9	US-09-747-835A-21	Sequence 21, Appli
42	75.5	16.1	4041	18	US-10-312-312-21	Sequence 21, Appli
43	75.5	16.1	4353	15	US-10-225-567A-405	Sequence 405, App
44	75.5	16.1	4755	21	US-10-505-486-201	Sequence 201, App
45	75.5	16.1	4914	9	US-09-974-298-188	Sequence 188, App

ALIGNMENTS

RESULT 1
US-09-917-378-2
; Sequence 2, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8


```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2289
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
US-09-917-378-2

Alignment Scores:
Pred. No.: 5,27e-56 Length: 2289
Score: 469.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-378-2 (1-2289)

Qy 1 ValSerGlyGlyValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTGAGGTGCAGTACAAACAATGATTCGGCGCGGTGATAACCCAG 1422
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTTCGACGGTG 1482
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGGTGGGGTACTGGTTCACCGGGATGGTGGGTCTGTCGACACTGGTGTACAACCTGTGAC 1542
Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGCGCGCGATGGGGTGTGGGAATATCCGCCCTCGTTCGGCTCGGTGAACCCGGCGACG 1602
Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACTACCTGCAG 1626

RESULT 2
US-09-917-384-2
; Sequence 2, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3687
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-2

Alignment Scores:
Pred. No.: 2,66e-55 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.15% Indels: 0
DB: 10 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-384-2 (1-3687)

Qy 1 ValSerGlyGlyValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAGGTGCAGTACAAACAATGATTCGGCGCGGTGATAACCCAG 1809
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTTCGACGGTG 1869
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGGGGTACTGGTTCACCGGGATGGTGGGTCTGTCGACACTGGTGTACAACCTGTGAC 1929
Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCGATGGGGTGTGGGAATATCCGCCCTCGTTCGGCTCGGTGAACCCGGCGACG 1989
Qy 81 ProThrAlaAspThrTyrLeuGln 88

US-09-917-376-4 (1-89) x US-09-917-383-2 (1-3687)

Qy 1 ValSerGlyGlyValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTGAGGTGCAGTACAAACAATGATTCGGCGCGGTGATAACCCAG 1809
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGGTCTGTCGGTGGATTTCGACGGTG 1869
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGGGGTACTGGTTCACCGGGATGGTGGGTCTGTCGACACTGGTGTACAACCTGTGAC 1929
Qy 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCGATGGGGTGTGGGAATATCCGCCCTCGTTCGGCTCGGTGAACCCGGCGACG 1989
Qy 81 ProThrAlaAspThrTyrLeuGln 88
```



```
Db 1990 CCGACGGCGGACACCTACCTGCAG 2013
|||||
RESULT 4
US-09-917-376-2
; Sequence 2, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2
Alignment Scores:
Pred. No.: 7,09e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 15 Gaps: 0
US-09-917-376-4 (1-89) x US-10-155-400-2 (1-2869)
QY 1 ValSerGlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAGGTGCAGTATAGATATATGATTCGGCGCGGGTGATAATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGTTTCAGGTGGTGAATACCGGGTCGTCGTCGGTGGATTGTGCGACGGTG 2724
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGGTTTCACCGGGATGGTGGCTCGTCGACACTGGTGTACAACTGTGAC 2784
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGATCGGGTGTGGGAATATCCGCGCTCGTTCGGTCTCGGTGAACCCGCGCAGC 2844
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868
RESULT 6
US-10-156-761-550
; Sequence 550, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 550
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) .. (2223)
US-10-156-761-550
Alignment Scores:
Pred. No.: 7,09e-55 Length: 2223
Score: 462.00 Matches: 86
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 97.73% Mismatches: 0
Query Match: 98.30% Indels: 0
DB: 15 Gaps: 0
US-09-917-376-4 (1-89) x US-09-917-376-2 (1-2869)
QY 1 ValSerGlyValIysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGTGAGGTGCAGTATAGATATATGATTCGGCGCGGGTGATAATCAG 2664
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGTTTCAGGTGGTGAATACCGGGTCGTCGTCGGTGGATTGTGCGACGGTG 2724
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGCGGTACTGGTTTCACCGGGATGGTGGCTCGTCGACACTGGTGTACAACTGTGAC 2784
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGCGCGCGATCGGGTGTGGGAATATCCGCGCTCGTTCGGTCTCGGTGAACCCGCGCAGC 2844
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACCTACCTGCAG 2868
RESULT 5
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
```


Pred. No.: 8.85e-26 Length: 2223
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.68% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-4 (1-89) x US-10-156-761-550 (1-2223)

Qy 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 1771 TCGGGCGGCTTCAAGGTCTCTACAAGAACACACACTCTCGGCCACCGAACAGCCCATC 1830

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 1831 CGGCCAGCCCTCGGATGTCACACCGGAGCGGCTCCCTCGACCTGTCCAAAGGTACAG 1890

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61

Db 1891 GCCCGCTACTACTTCAACCGGACAGCGCTCGCCACCGTGAACGGCTGGTGGCGACTAC 1950

Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 1951 GCGCGCGTCTGTTCCACGTGAGCTGAGCTGAGTGGTACCCCTGACACCGCCGTCGCG 2010

Qy 82 ThrAlaAspThrTyrLeuGln 88

Db 2011 GGAGCGGAGCTTACTCTCGAA 2031

RESULT 7

US-10-156-761-1

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 1

; LENGTH: 9025608

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Alignment Scores:

Pred. No.: 5.62e-21 Length: 9025608

Score: 257.00 Matches: 49

Percent Similarity: 71.26% Conservative: 13

Best Local Similarity: 56.32% Mismatches: 25

Query Match: 54.68% Indels: 0

DB: 15 Gaps: 0

US-09-917-376-4 (1-89) x US-10-156-761-1 (1-9025608)

Qy 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21

Db 706216 TCGGGCGGCTTCAAGGTCTCTACAAGAACACACACTCTCGGCCACCGAACAGCCCATC 706275

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 706276 CGGCCAGCCCTCGGATGTCACACCGGAGCGCTCCCTCGACCTGTCCAAGGTACAG 706335

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61
Db 706336 GCCCGCTACTACTTCAACCGGACAGCGCTCGCCACCGTGAACGCCCTGGTGGCGACTAC 706395

Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 706396 GCGCGCGTCTGTTCCACGTGAGCTGAGTGGTACCCCTGACACCGCCGTCGCG 706455

Qy 82 ThrAlaAspThrTyrLeuGln 88
Db 706456 GGAGCGGAGCTTACTCTCGAA 706476

RESULT 8

US-10-369-493-46838

; Sequence 46838, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xiaofeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 46838

; LENGTH: 1527

; TYPE: DNA

; ORGANISM: Bacillus subtilis

US-10-369-493-46838

Alignment Scores:

Pred. No.: 1.94e-16 Length: 1527

Score: 190.00 Matches: 35

Percent Similarity: 65.12% Conservative: 21

Best Local Similarity: 40.70% Mismatches: 26

Query Match: 40.43% Indels: 4

DB: 17 Gaps: 2

US-09-917-376-4 (1-89) x US-10-369-493-46838 (1-1527)

Qy 4 GlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 1084 GGTATTTCTGTACAGTACAGCAGCGGATGGAGTATGAACAGCAACCAATCCGTCGG 1143

Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1144 CAGCTTCAATATAAAATAACGCAATACCGGTTGATTATAAGATGTCTACGCCGT 1203

Qy 44 TyrTrpPhe---ThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrpAla 62
Db 1204 TACTGGTATAAAGCGAAACAAAGGCCCAAC-----TTTGACTGTGACTACGCG 1254

Qy 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82
Db 1255 CAGATTGGATCGGCAATGTGACACACAAGTTTGTGACGTTGACATAAACAAGCAAGGT 1314

Qy 83 AlaAspThrTyrLeuGln 88

Db 1315 GCAGATACCTATCTCGAA 1332

RESULT 9

US-10-466-208-11


```
; Sequence 11, Application US/10466208
; Publication No. US20040180348A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 11
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCCase gene with additional his encoding sequences
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1431)
US-10-466-208-11
Alignment Scores:
Pred. No.: 3,44e-16 Length: 1434
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservativeness: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 19 Gaps: 1
US-09-917-376-4 (1-89) x US-10-466-208-11 (1-1434)
Qy 4 GlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 991 GGTATTCTGTACAAATACAGAGCGGGGATGGAGTATGAACAGCAACCAATCCGTCGG 1050
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1051 CAGCTTCAATATAAATAAATACGCAATACCAAGCGTTGATTAAAGAGATGTCACTGCCCGT 1110
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCyAspTrpAlaAla 63
Db 1111 TACTGGTATACGCGAATAAACAAGCGCAAAACGTT-----GACTGTGACTACGCGCAG 1164
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1165 CTTGGATGCGCAATGTGCATACACAAGTTTGTGACGTTGTCATTAACCAACCAAGCAAGTGCA 1224
Qy 84 AspThrTyrLeuGln 88
Db 1225 GATACCTATCTGGAA 1239
RESULT 10
US-10-466-208-7
; Sequence 7, Application US/10466208
; Publication No. US20040180348A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 7
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCCase gene with mutated signal sequence to enhance its
```

```
; OTHER INFORMATION: hydrophobicity
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1488)
US-10-466-208-7
Alignment Scores:
Pred. No.: 3,62e-16 Length: 1491
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservativeness: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 19 Gaps: 1
US-09-917-376-4 (1-89) x US-10-466-208-7 (1-1491)
Qy 4 GlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 1048 GGTATTCTGTACAAATACAGAGCGGGGATGGAGTATGAACAGCAACCAATCCGTCGG 1107
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1108 CAGCTTCAATATAAATAAATACGCAATACCAAGCGTTGATTAAAGAGATGTCACTGCCCGT 1167
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCyAspTrpAlaAla 63
Db 1168 TACTGGTATACGCGAATAAACAAGCGCAAAACGTT-----GACTGTGACTACGCGCAG 1221
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1222 CTTGGATGCGCAATGTGCATACACAAGTTTGTGACGTTGTCATTAACCAACCAAGCAAGTGCA 1281
Qy 84 AspThrTyrLeuGln 88
Db 1282 GATACCTATCTGGAA 1296
RESULT 11
US-10-433-577-34
; Sequence 34, Application US/10433577
; Publication No. US20040171065A1
; GENERAL INFORMATION:
; APPLICANT: Genofocus Co., Ltd.
; TITLE OF INVENTION: Method for Expression of Proteins on Spore Surface
; FILE REFERENCE: PCT-Genofocus-1
; CURRENT APPLICATION NUMBER: US/10/433,577
; CURRENT FILING DATE: 2003-06-05
; PRIOR APPLICATION NUMBER: KR2000-74835
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 34
; LENGTH: 2510
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (461)..(2491)
US-10-433-577-34
Alignment Scores:
Pred. No.: 7,25e-16 Length: 2510
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservativeness: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.00% Indels: 2
DB: 19 Gaps: 1
US-09-917-376-4 (1-89) x US-10-433-577-34 (1-2510)
Qy 4 GlyValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2051 GGTATTCTGTACAAATACAGAGCGGGGATGGAGTATGAACAGCAACCAATCCGTCGG 2110
```



```
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 2111 CAGCTTCABATAAAATACGGCAATACACCGTGTGATTAAAGATGTCACCTGCCCGT 2170
Qy 44 TyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 2171 TACTGGTATACCGCAAAACCAAGGCAAAACCGTT-----GACTGTGACTACGCGCAG 2224
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2225 CTTGGATGCGCAATGTTGACATACAGTTTGTGACGTTTGCAATAACCAAGCAAGGTGCA 2284
Qy 84 AspThrTyrLeuGln 88
Db 2285 GATACCTATCTGGAA 2299

RESULT 12
US-10-933-404-1
; Sequence 1, Application US/10933404
; Publication No. US20050118729A1
; GENERAL INFORMATION:
; APPLICANT: MORAG, Ely
; TITLE OF INVENTION: MICROARRAYS OF CELLULOSE BINDING CHIMERIC PROTEINS AND METHODS OF
; FILE REFERENCE: 85189-7100
; CURRENT APPLICATION NUMBER: US/10/933,404
; CURRENT FILING DATE: 2004-09-03
; PRIOR APPLICATION NUMBER: PCT/IL03/00177
; PRIOR FILING DATE: 2003-03-06
; PRIOR APPLICATION NUMBER: US 60/362,061
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-933-404-1

Alignment Scores:
Pred. No.: 1.62e-13 Length: 501
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 21

US-09-917-376-4 (1-89) x US-10-933-404-1 (1-501)

Qy 1 ValSerGlyGlyValIleValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 13 GTATCAGGCAATTTGAAGTTGAATCTTACACAGCAATCTCTCAGATACCTACTACTCA 72
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 73 ATCAATCTCAGTTCAGTTACTAATACCGAGCAGTGCATGTAATGATTGTCCAAATC 132
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 133 ACATTGAGATATTATTATACAGTAGACGAGCAGAAAGATCAGACCTTCTGG--TGTGAC 189
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 190 CATGCTGCAATAATCGGCAGTACGAGCTACACCGGAATTAATCTCAATGTAAAGGA 249
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 250 ACATTTGTAAATAAGTTCCTCAACAAATAACGACGACACCTACCTTGAA 300

RESULT 13
US-10-372-054-3
; Sequence 3, Application US/10372054
; Publication No. US20030203466A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. i633
US-10-372-054-3

Alignment Scores:
Pred. No.: 6.59e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.11% Indels: 10
DB: 17

US-09-917-376-4 (1-89) x US-10-372-054-3 (1-1438)

Qy 1 ValSerGlyGlyValIleValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTTGAAGTTGAATCTTACACAGCAATCTCTCAGATACCTACTACTCA 999
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCTCAGTTCAGTTACTAATACCGAAGCAGTGCATGTAATGATTGTCCAAATC 1059
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTAGACGAGCAGAAAGATCAGACCTTCTGG--TGTGAC 1116
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATAATCGGCAGTACGAGCTACACCGGAATTAATCTCAATGTAAAGGA 1176
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTTGTAAATAAGTTCCTCAACAAATAACGACGACACCTACCTTGAA 1227

RESULT 14
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
```



```
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-655-433-11

Alignment Scores:
Pred. No.:      6,86e-13      Length:      1482
Score:          165.00      Matches:      35
Percent Similarity: 59.79%      Conservative: 23
Best Local Similarity: 36.08%      Mismatches: 29
Query Match:     35.11%      Indels:      10
DB:              18          Gaps:          2

US-09-917-376-4 (1-89) x US-10-655-433-11 (1-1482)
QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTCAGGTTGAATCTTACACAGCAATCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTTACTAATACCGAAGCAGTGCAATTGATTTGTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATTAATCGGCAGTAAACGGCAGCTACAACGGAAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAAATAACGCAGACACCTACCTTGAA 1278

Search completed: October 6, 2005, 04:57:57
Job time : 911.446 secs
```

```
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.:      6,86e-13      Length:      1482
Score:          165.00      Matches:      35
Percent Similarity: 59.79%      Conservative: 23
Best Local Similarity: 36.08%      Mismatches: 29
Query Match:     35.11%      Indels:      10
DB:              13          Gaps:          2

US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)
QY 1 ValSerGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGCAATTTCAGGTTGAATCTTACACAGCAATCTTCAGATACTACTAACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTTACTAATACCGAAGCAGTGCAATTGATTTGTCCAAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATTAATCGGCAGTAAACGGCAGCTACAACGGAAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGTAATAATGAGTTCTCTCAACAAATAACGCAGACACCTACCTTGAA 1278
```

```
RESULT 15
US-10-655-433-11
; Sequence 11, Application US/10655433
; Publication No. US20040067572A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20040067572A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/655,433
; PRIOR FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:48:48 ; Search time 16.8581 Seconds
(without alignments)
2018.911 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKQVKNDSAPGDNQ.....IRASFGSVNPATPDYTLQ 88

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04.*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	100.0	88	7 ADD22925	Add22925 Acidother
2	469	100.0	89	7 ADD22924	Add22924 Acidother
3	469	100.0	154	6 ABP73020	Abp73020 Amino aci
4	469	100.0	154	7 ADJ38294	Adj38294 A. cellul
5	469	100.0	762	6 ABP73022	Abp73022 Amino aci
6	469	100.0	762	7 ADJ38291	Adj38291 A. cellul
7	466	99.4	150	6 ABP73025	Abp73025 Amino aci
8	466	99.4	150	8 ADH36640	Adh36640 A. cellul
9	466	99.4	1228	6 ABP73029	Abp73029 Amino aci
10	466	99.4	1228	8 ADH36636	Adh36636 Acidother
11	462	98.5	88	6 ABP73018	Abp73018 Amino aci
12	462	98.5	88	6 AD52317	Ad52317 A. cellul
13	462	98.5	89	6 ABP73017	Abp73017 Amino aci
14	462	98.5	89	8 AD52316	Ad52316 A. cellul
15	462	98.5	957	6 ABP73015	Abp73015 Amino aci
16	462	98.5	957	7 ADP22921	Adp22921 Acidother
17	462	98.5	957	8 AD52313	Ad52313 Acidother
18	436	93.0	153	6 ABP71658	Abp71658 A. cellul
19	436	93.0	153	7 ADP75898	Adp75898 Carbohyd
20	436	93.0	1121	6 ABP71656	Abp71656 A. cellul
21	436	93.0	1121	7 ADP75895	Adp75895 Acidother
22	231.5	49.4	782	2 AAR15625	Aar15625 Cellulase
23	195.5	41.7	616	2 AAY13494	Aay13494 Truncated
24	195.5	41.7	616	5 AAE16325	Aae16325 Active ce
25	194.5	41.5	1751	2 AAY13493	Aay13493 Truncated

ALIGNMENTS

RESULT 1
ADD22925
ID ADD22925 standard; protein; 88 AA.

XX AC ADD22925;
XX DT 15-JAN-2004 (first entry)
XX DE Acidothermus cellulolyticus avicelase AvIII CBD III #2.
XX KW enzyme; AvIII; cellulose reduction; agricultural biomass;
XX KW municipal solid waste; glycoside hydrolase; avicelase.
XX OS Acidothermus cellulolyticus.
XX PN US2003108988-A1.
XX PD 12-JUN-2003.
XX PF 18-OCT-2002; 2002US-00155400.
XX PR 28-JUL-2001; 2001US-00917376.
XX PA (DING/) DING S.
XX PA (ADNE/) ADNEY W S.
XX PA (VINZ/) VINZANT T B.
XX PI (HIMM/) HIMMEL M E.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-810853/76.
XX New isolated thermal tolerant avicelase polynucleotide useful for
XX detection of a polynucleotide encoding AvIII and for reducing cellulose
XX in a starting material, e.g. municipal solid waste.
XX Claim 16; SEQ ID NO 5; 29pp; English.
XX The invention relates to an isolated polynucleotide molecule encoding a
XX thermostable AvIII polypeptide. The polynucleotide is useful for
XX detection of a polynucleotide encoding AvIII. The polynucleotide is
XX useful for reducing cellulose in a starting material which involves
XX administering to the starting material, e.g. agricultural biomass or
XX municipal solid waste, a polypeptide molecule of the polynucleotide. The
XX method further comprises administering a second polypeptide molecule
XX chosen from the glycoside hydrolase family of proteins. The present
XX sequence represents the amino acid sequence of Acidothermus
XX cellulolyticus avicelase AvIII CBD III #2.

26 194.5 41.5 1751 5 AAE16324
27 191.5 40.8 1426 2 AAY13492
28 191.5 40.8 1426 5 AAE16323
29 190 40.5 508 7 ADO55907
30 190 40.5 508 8 ADS44721
31 188 40.1 477 5 AAO22444
32 188 40.1 496 5 AAO22443
33 188 40.1 677 5 AAU98063
34 186 39.7 499 2 AAR42122
35 172 36.7 700 2 AAR13227
36 165.5 35.3 551 2 AAW18790
37 165 35.2 167 2 AAR95080
38 165 35.2 167 7 ADG14244
39 165 35.2 239 7 ADG14258
40 165 35.2 476 3 AAY54123
41 165 35.2 493 2 AAY28850
42 165 35.2 493 2 AAY43218
43 165 35.2 531 2 AAW15238
44 165 35.2 1853 2 AAW43108
45 163 34.8 499 8 ADJ35112

Aae16324 Active ce
Aay13492 Truncated
Aae16323 Active ce
Aao55907 Bacillus
Ada44721 Bacterial
Aao22444 Protein e
Aao22443 Protein e
Aau98063 Bacillus
Aar42122 NK-1 cell
Aar13227 Novel end
Aaw18790 Corrected
Aar95080 Cellulose
Adg14244 Cellulose
Adg14258 Cellulose
Aay54123 A mannana
Aay28850 Pectate 1
Aay43218 Pectate 1
Aaw15238 Scaffold
Aaw43108 C. thermo
Adj35112 Xylanase

XX SQ Sequence 88 AA;

Query Match 100.0%; Score 469; DB 7; Length 88;
 Best Local Similarity 100.0%; Pred. No. 5.4e-46;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60
 DB 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60

OY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 2
 ADD22924
 ID ADD22924 standard; protein; 89 AA.
 XX
 AC ADD22924;
 DT 15-JAN-2004 (first entry)
 DE
 DE Acidothermus cellulolyticus avicelase AvIII CBD III #1.
 XX
 KW enzyme; AvIII; cellulose reduction; agricultural biomass;
 KW municipal solid waste; glycoside hydrolase; avicelase.
 XX
 OS Acidothermus cellulolyticus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 89 /label= Unknown
 FT
 XX
 XX US2003108988-A1.
 XX
 PD 12-JUN-2003.
 XX
 XX 18-OCT-2002; 2002US-00155400.
 PF
 XX 28-JUL-2001; 2001US-00917376.
 PR
 XX (DING/) DING S.
 PA (ADNE/) ADNEY W S.
 PA (VINZ/) VINZANT T B.
 PA (HIMM/) HIMMEL M E.
 XX
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;
 XX WPI; 2003-810853/76.
 DR
 XX New isolated thermal tolerant avicelase polynucleotide useful for
 PT detection of a polynucleotide encoding AvIII and for reducing cellulose
 PT in a starting material, e.g. municipal solid waste.
 XX
 PS Claim 16; SEQ ID NO 4; 29pp; English.
 XX
 CC The invention relates to an isolated polynucleotide molecule encoding a
 CC thermostable AvIII polypeptide. The polynucleotide is useful for
 CC detection of a polynucleotide encoding AvIII. The polynucleotide is
 CC useful for reducing cellulose in a starting material which involves
 CC administering to the starting material, e.g. agricultural biomass or
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
 CC method further comprises administering a second polypeptide molecule
 CC chosen from the glycoside hydrolase family of proteins. The present
 CC sequence represents the amino acid sequence of Acidothermus
 CC cellulolyticus avicelase AvIII CBD III #1.
 XX
 SQ Sequence 89 AA;

Query Match 100.0%; Score 469; DB 7; Length 89;
 Best Local Similarity 100.0%; Pred. No. 5.5e-46;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60
 DB 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60

OY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 3
 ABP73020
 ID ABP73020 standard; peptide; 154 AA.
 XX
 AC ABP73020;
 XX
 DT 03-JUN-2003 (first entry)
 DE
 DE Amino acid sequence of the ManA carbohydrate binding domain type III.
 XX
 KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
 KW food; feed; paper pulp; biofuel; mannase.
 XX
 OS Acidothermus cellulolyticus.
 XX
 PN WO2003012110-A1.
 XX
 PD 13-FEB-2003.
 XX
 XX 28-JUL-2001; 2001WO-US023819.
 PF
 XX 28-JUL-2001; 2001WO-US023819.
 PR
 XX (MIDE) MIDWEST RES INST.
 PA
 XX Ding S, Adney WS, Vinzant TB, Himmel ME;
 PI WPI; 2003-248182/24.
 XX
 XX Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
 PT cellulolyticus, useful for reducing hemicellulose in a starting material,
 PT for processing of food, and as bulking agents in food stuffs.
 XX
 PS Claim 8; Page 7; 46pp; English.
 XX
 CC The present sequence represents a fragment of ManA, a thermostable
 CC mannanase A polypeptide derived from Acidothermus cellulolyticus. ManA is
 CC a member of the glycoside hydrolase family of enzymes. ManA is useful for
 CC reducing hemicellulose in a starting material to simpler carbohydrate
 CC units, and ultimately to sugars which are useful in the food, feed, paper
 CC pulp, and biofuels industries. It is useful for the processing of food
 CC and in food stuffs as bulking agents, and for the degradation of mannan.
 CC ManA is also useful to raise polyclonal and monoclonal antibodies that
 CC are useful in purifying ManA, or detecting ManA polypeptide expression,
 CC and as well as reagent tools for characterizing the molecular actions of
 CC ManA polypeptides
 XX
 SQ Sequence 154 AA;

Query Match 100.0%; Score 469; DB 6; Length 154;
 Best Local Similarity 100.0%; Pred. No. 1.1e-45;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60
 DB 1 VSGGVKVOYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRWFTRDGGSTLVNCD 60

OY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88


```
RESULT 4
ADJ38294
ID ADJ38294 standard; protein; 154 AA.
XX AC ADJ38294;
XX DT 06-MAY-2004 (first entry)
XX DE A. cellulolyticus mannanase ManA carbohydrate binding domain III.
XX KW mannanase A; catalytic domain GH5; carbohydrate binding domain III;
XX KW carbohydrate binding domain II; food processing; foodstuff;
XX KW bulking agent; ManA; enzyme; EC3.2.1.78.
XX OS Acidothermus cellulolyticus.
XX PN US2003119093-A1.
XX PD 26-JUN-2003.
XX PF 28-JUL-2001; 2001US-00917378.
XX PR 28-JUL-2001; 2001US-00917378.
XX PA (DING/) DING S.
XX PA (ADNEY/) ADNEY W S.
XX PA (VINZ/) VINZANT T B.
XX PA (HIMM/) HIMMEL M E.
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-863435/80.
XX New composition comprising a purified mannanase A peptide, comprising a
XX catalytic domain GH5, carbohydrate binding domain III or carbohydrate
XX binding domain II, useful in food processing and in food stuffs as
XX bulking agents.
XX Claim 7; SEQ ID NO 4; 18pp; English.
XX The invention relates to a composition comprising a purified mannanase A
XX peptide, having a catalytic domain GH5, carbohydrate binding domain III
XX or carbohydrate binding domain II. The composition is useful in food
XX processing and in foodstuffs as bulking agents. The present sequence
XX represents the amino acid sequence of A. cellulolyticus mannanase ManA
XX carbohydrate binding domain III.
XX Sequence 154 AA;
Query Match 100.0%; Score 469; DB 7; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.1e-45;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 60
Db 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 60
QY 61 WAAMGCGNIRASFGSVNPTPTADTYLQ 88
Db 61 WAAMGCGNIRASFGSVNPTPTADTYLQ 88
RESULT 5
ABP73022
ID ABP73022 standard; protein; 762 AA.
XX AC ABP73022;
XX DT 03-JUN-2003 (first entry)
XX DE Amino acid sequence of the ManA polypeptide.
XX ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
```

```
XX food; feed; paper pulp; biofuel; mannanase.
XX Acidothermus cellulolyticus.
XX WO2003012110-A1.
XX 13-FEB-2003.
XX 28-JUL-2001; 2001WO-US023819.
XX 28-JUL-2001; 2001WO-US023819.
XX (MIDE ) MIDWEST RES INST.
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
XX WPI; 2003-248182/24.
XX N-PSDB; ABZ77633.
XX Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
XX cellulolyticus, useful for reducing hemicellulose in a starting material,
XX for processing of food, and as bulking agents in food stuffs.
XX Claim 3; Page 18-19; 46pp; English.
XX The present sequence represents ManA, a thermostable mannanase A
XX polypeptide derived from Acidothermus cellulolyticus. ManA is a member of
XX the glycoside hydrolase family of enzymes. ManA is useful for reducing
XX hemicellulose in a starting material to simpler carbohydrate units, and
XX ultimately to sugars which are useful in the food, feed, paper pulp, and
XX biofuels industries. It is useful for the processing of food and in food
XX stuffs as bulking agents, and for the degradation of mannanase. ManA is
XX also useful to raise polyclonal and monoclonal antibodies that are useful
XX in purifying ManA, or detecting ManA polypeptide expression, and as well
XX as reagent tools for characterizing the molecular actions of ManA
XX polypeptides
XX Sequence 762 AA;
Query Match 100.0%; Score 469; DB 6; Length 762;
Best Local Similarity 100.0%; Pred. No. 7.8e-45;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 60
Db 455 VSGGVKQVQKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLYVNC 514
QY 61 WAAMGCGNIRASFGSVNPTPTADTYLQ 88
Db 515 WAAMGCGNIRASFGSVNPTPTADTYLQ 542
RESULT 6
ADJ38291
ID ADJ38291 standard; protein; 762 AA.
XX AC ADJ38291;
XX DT 06-MAY-2004 (first entry)
XX DE A. cellulolyticus mannanase ManA.
XX mannanase A; catalytic domain GH5; carbohydrate binding domain III;
XX carbohydrate binding domain II; food processing; foodstuff;
XX bulking agent; ManA; enzyme; EC3.2.1.78.
XX OS Acidothermus cellulolyticus.
XX PN US2003119093-A1.
XX PD 26-JUN-2003.
XX 28-JUL-2001; 2001US-00917378.
```



```
XX 28-JUL-2001; 2001US-00917378.
PR (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
DR WPI; 2003-863435/80.
DR N-ESDB; ADJ38292.
XX
XX New composition comprising a purified mannanase A peptide, comprising a
PT catalytic domain GH5, carbohydrate binding domain III or carbohydrate
PT binding domain II, useful in food processing and in food stuffs as
PT bulking agents.
XX
XX Claim 10; SEQ ID NO 1; 18pp; English.
PS
CC The invention relates to a composition comprising a purified mannanase A
CC peptide, having a catalytic domain GH5, carbohydrate binding domain III
CC or carbohydrate binding domain II. The composition is useful in food
CC processing and in foodstuffs as bulking agents. The present sequence
CC represents the amino acid sequence of A. cellulolyticus mannanase ManA.
XX
XX Sequence 762 AA;
SQ
Query Match 100.0%; Score 469; DB 7; Length 762;
Best Local Similarity 100.0%; Pred. No. 7.8e-45;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSSSTLVNCD 60
DB 455 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSSSTLVNCD 514
OY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542
XX
RESULT 7
ABP73025
ID ABP73025 standard; peptide; 150 AA.
XX
AC ABP73025;
XX
DT 03-JUN-2003 (first entry)
XX
DE Amino acid sequence of the GuxA carbohydrate binding domain type III.
XX
KW GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile.
XX
OS Acidothermus cellulolyticus.
XX
XX WO2003012109-A1.
PN
XX 13-FEB-2003.
PD
XX 28-JUL-2001; 2001WO-US023817.
PF
XX 28-JUL-2001; 2001WO-US023817.
PR
XX (MIDE ) MIDWEST RES INST.
PA
XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
PI
XX WPI; 2003-239526/23.
DR
XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing cellulose in a starting material, and
PT for the conversion of biomass to biofuels and biofuel additives.
XX
XX Claim 27; SEQ ID NO 5; 20pp; English.
PS
XX The present invention relates to the isolation of a thermal tolerant
CC cellulase from Acidothermus cellulolyticus. The cellulase is a member of
CC
```

```
XX Claim 6; Page 7; 47pp; English.
PS
CC ABP73023-28 represent peptide fragments of a GuxA polypeptide. GuxA is a
CC thermostable cellulase, and is a member of the glycoside hydrolase family
CC of enzymes. GuxA is useful for reducing cellulose in a starting material
CC such as agricultural biomass to sugars. This is useful in biofuel
CC production. GuxA is also useful in the conversion of biomass to biofuels
CC and biofuel additives, in detergents, pulp and paper processing, food and
CC feed processing, and in textile process. GuxA is also useful for raising
CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,
CC or detecting GuxA polypeptide expression, as well as reagent tools for
CC characterizing the molecular actions of GuxA polypeptides
XX
XX Sequence 150 AA;
SQ
Query Match 99.4%; Score 466; DB 6; Length 150;
Best Local Similarity 98.9%; Pred. No. 2.3e-45;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 VSGGVKVOYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSSSTLVNCD 60
DB 1 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTRDGSSSTLVNCD 60
OY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
XX
RESULT 8
ADH36640
ID ADH36640 standard; protein; 150 AA.
XX
AC ADH36640;
XX
DT 11-MAR-2004 (first entry)
XX
DE A. cellulolyticus glycoside hydrolase, GuxA CBD type III domain.
XX
KW Thermal tolerant cellulase; glycoside hydrolase; GuxA;
KW carbohydrate degradation; cellulose; agricultural biomass;
KW municipal solid waste; thermostable; enzyme;
KW carbohydrate binding domain type III; CBD III.
XX
OS Acidothermus cellulolyticus.
XX
XX US2003104522-A1.
PN
XX 05-JUN-2003.
PD
XX 28-JUL-2001; 2001US-00917383.
PF
XX 28-JUL-2001; 2001US-00917383.
PR
XX (DING/) DING S.
PA
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
PA (DECK/) DECKER S R.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX
DR WPI; 2004-106451/11.
DR
XX Novel isolated thermostable GuxA polypeptide useful for detecting
PT polynucleotide encoding GuxA, assessing carbohydrate degradation activity
PT of GuxA, reducing cellulose in starting material e.g., agricultural
PT biomass.
XX
XX Claim 27; SEQ ID NO 5; 20pp; English.
PS
XX The present invention relates to the isolation of a thermal tolerant
CC cellulase from Acidothermus cellulolyticus. The cellulase is a member of
CC
```


CC the glycoside hydrolase family and is designated GuxA. Also disclosed are
 CC methods of producing and using GuxA. The GuxA polypeptide is useful for
 CC the detection of a polynucleotide encoding GuxA. The polypeptide sequence
 CC is also useful for assessing the carbohydrate, e.g. cellulose,
 CC degradation activity of GuxA. GuxA is useful for reducing cellulose in a
 CC starting material such as agricultural biomass or municipal solid waste.
 CC The polypeptide molecule of GuxA is thermostable. The present sequence
 CC represents a domain of the A. cellulolyticus GuxA polypeptide.
 XX
 SQ Sequence 150 AA;

Query Match 99.4%; Score 466; DB 8; Length 150;
 Best Local Similarity 98.9%; Pred. No. 2.3e-45;
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSTLVYVNC 60
 DB 1 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSTLVYVNC 60
 QY 61 WAAMGCGNIRASFGSVNPTPTADTYLQ 88
 DB 61 WAAMGCGNIRASFGSVNPTPTADTYLQ 88

RESULT 9
 ABP73029
 ID ABP73029 standard; protein; 1228 AA.

XX
 AC ABP73029;

DT 03-JUN-2003 (first entry)

XX Amino acid sequence of the GuxA polypeptide.

XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
 KW detergent; pulp processing; paper processing; feed processing; textile.

XX Acidothermus cellulolyticus.

XX WO2003012109-A1.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023817.

XX 28-JUL-2001; 2001WO-US023817.

XX (MIDE) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2003-239526/23.

XX N-PSDB; ABZ77634.

XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus
 PT cellulolyticus, useful for reducing cellulose in a starting material, and
 PT for the conversion of biomass to biofuels and biofuel additives.

XX Claim 3; Page 19; 47pp; English.

XX The present sequence represents a GuxA polypeptide. GuxA is thermostable
 CC cellulase, and is a member of the glycoside hydrolase family of enzymes.
 CC GuxA is useful for reducing cellulose in a starting material such as
 CC agricultural biomass to sugars. This is useful in biofuel production.
 CC GuxA is also useful in the conversion of biomass to biofuels and biofuel
 CC additives, in detergents, pulp and paper processing, food and feed
 CC processing, and in textile process. GuxA is also useful for raising
 CC polyclonal and monoclonal antibodies that are useful in purifying GuxA,
 CC or detecting GuxA polypeptide expression, as well as reagent tools for
 CC characterizing the molecular actions of GuxA polypeptides

XX Sequence 1228 AA;

Query Match 99.4%; Score 466; DB 6; Length 1228;
 Best Local Similarity 98.9%; Pred. No. 3.1e-44;
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSTLVYVNC 60
 DB 584 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSTLVYVNC 643
 QY 61 WAAMGCGNIRASFGSVNPTPTADTYLQ 88
 DB 644 WAAMGCGNIRASFGSVNPTPTADTYLQ 671

RESULT 10
 ADH36636

ID ADH36636 standard; protein; 1228 AA.

XX
 AC ADH36636;

XX 11-MAR-2004 (first entry)

XX Acidothermus cellulolyticus glycoside hydrolase, GuxA.

XX Thermal tolerant cellulase; glycoside hydrolase; GuxA;
 KW carbohydrate degradation; cellulose; agricultural biomass;
 KW municipal solid waste; thermostable; enzyme.

XX Acidothermus cellulolyticus.

XX US2003104522-A1.

XX 05-JUN-2003.

XX 28-JUL-2001; 2001US-00917383.

XX 28-JUL-2001; 2001US-00917383.

XX (DING/) DING S.

XX (ADNE/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX (DECK/) DECKER S R.

XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;

XX WPI; 2004-106451/11.

XX N-PSDB; ADH36637.

XX Novel isolated thermostable GuxA polypeptide useful for detecting
 PT polynucleotide encoding GuxA, assessing carbohydrate degradation activity
 PT of GuxA, reducing cellulose in starting material e.g., agricultural
 PT biomass.

XX Claim 27; SEQ ID NO 1; 20pp; English.

XX The present invention relates to the isolation of a thermal tolerant
 CC cellulase from Acidothermus cellulolyticus. The cellulase is a member of
 CC the glycoside hydrolase family and is designated GuxA. Also disclosed are
 CC methods of producing and using GuxA. The GuxA polypeptide is useful for
 CC the detection of a polynucleotide encoding GuxA. The polypeptide sequence
 CC is also useful for assessing the carbohydrate, e.g. cellulose,
 CC degradation activity of GuxA. GuxA is useful for reducing cellulose in a
 CC starting material such as agricultural biomass or municipal solid waste.
 CC The polypeptide molecule of GuxA is thermostable. The present sequence
 CC represents A. cellulolyticus glycoside hydrolase, GuxA.

XX Sequence 1228 AA;

Query Match 99.4%; Score 466; DB 8; Length 1228;
 Best Local Similarity 98.9%; Pred. No. 3.1e-44;
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYVWFTTRDGGSTLVYVNC 60


```
||||:|||||
584 VSGGLKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 643
Qy 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 644 WAAMGCGNIRASFGSVNPATPTADTYLQ 671

RESULT 11
ABP73018
ID ABP73018 standard; peptide; 88 AA.
AC ABP73018;
XX
XX 03-JUN-2003 (first entry)
XX
XX Amino acid sequence of AvIII carbohydrate binding domain type II.
DE
XX
XX Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;
KW detergent; pulp processing; paper processing; feed processing; textile;
KW cellulose.
XX
XX Acidothermus cellulolyticus.
OS
XX
XX WO2003012090-A2.
XX
XX 13-FEB-2003.
XX
XX 28-JUL-2001; 2001WO-US023818.
XX
XX 28-JUL-2001; 2001WO-US023818.
XX
XX (MIDE ) MIDWEST RES INST.
XX
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
PI
XX WPI; 2003-248177/24.
XX
XX New thermostable AvIII peptide from Acidothermus cellulolyticus, useful
PT for degradation of cellulose or in generating anti-AvIII antibodies for
PT purifying recombinant AvIII polypeptides from genetically engineered
PT host cells.
XX
XX Claim 6; Page 8; 44pp; English.
XX
XX The present sequence is derived from a thermostable avicelase, designated
CC AvIII. AvIII is a member of the glycoside hydrolase family of enzymes,
CC and is a cellulase. AvIII is useful in the conversion of biomass to
CC biofuels and biofuel additives. It may be useful in the production of
CC detergents, pulp and paper processing, food and feed processing and in
CC textile processes. The thermostable AvIII peptide is useful in the
CC degradation of cellulose, and in generating specific anti-AvIII
CC antibodies that are useful in purifying recombinant AvIII polypeptides
CC from genetically engineered host cells, in detecting AvIII polypeptide
CC expression, as well as a reagent tool for characterizing the molecular
CC actions of the polypeptide. The AvIII polynucleotide is useful as a
CC source of probes or primers in various diagnostic assays
XX
XX Sequence 88 AA;
Query Match 98.5%; Score 462; DB 6; Length 88;
Best Local Similarity 97.7%; Pred. No. 3.5e-45;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 60
Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 60
Qy 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAIGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 12
AD052317
ID AD052317 standard; protein; 88 AA.
XX
XX AC AD052317;
XX
XX 15-JUL-2004 (first entry)
XX
XX A. cellulolyticus AvIII CBD_III partial protein #2.
XX
XX Thermostable cellulase; avicelase III; AvIII; cellulose reduction;
KW biomass degradation; ethanol formation; industrial chemical;
KW fabric treatment; carbohydrate binding domain type III; CBD III.
XX
XX Acidothermus cellulolyticus.
OS
XX
XX US2004038334-A1.
XX
XX 26-FEB-2004.
XX
XX 28-JUL-2001; 2001US-00917376.
XX
XX 28-JUL-2001; 2001US-00917376.
XX
XX (DING/) DING S.
XX (ADNE/) ADNEY W S.
XX (VINZ/) VINZANT T B.
XX (HIMM/) HIMMEL M E.
XX
XX Ding S, Adney WS, Vinzant TB, Himmel ME;
PI
XX WPI; 2004-203224/19.
XX
XX Novel thermostable AvIII polypeptide of glycoside hydrolase family and
PT isolated from Acidothermus cellulolyticus, useful for degrading treated
PT biomass into simpler forms of carbohydrate.
XX
XX Claim 28; SEQ ID NO 5; 19pp; English.
XX
XX The invention relates to a thermostable cellulase enzyme, avicelase III
CC (AvIII) and its nucleic acid sequence. AvIII is useful for reducing
CC cellulose in a starting material. A thermostable AvIII peptide is useful
CC for degrading treated biomass into simpler forms of carbohydrate, which
CC is used in the formation of ethanol or other industrial chemicals. It is
CC also useful for treating fabrics to remove cellulose-containing stains.
CC The present sequence is Acidothermus cellulolyticus AvIII carbohydrate
CC binding domain type III (CBD_III) partial protein.
XX
XX Sequence 88 AA;
Query Match 98.5%; Score 462; DB 8; Length 88;
Best Local Similarity 97.7%; Pred. No. 3.5e-45;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 60
Db 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGSSTLVNCD 60
Qy 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAIGCGNIRASFGSVNPATPTADTYLQ 88

RESULT 13
ABP73017
ID ABP73017 standard; peptide; 89 AA.
XX
XX AC ABP73017;
XX
XX 03-JUN-2003 (first entry)
XX
XX Amino acid sequence of AvIII carbohydrate binding domain type III.
DE
XX
```


KW Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;
 KW detergent; pulp processing; paper processing; feed processing; textile;
 KW cellulose.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers
 XX Misc-difference 89 /note= "unspecified residue"
 FT

XX WO2003012090-A2.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023818.

XX 28-JUL-2001; 2001WO-US023818.

XX (MIDE) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;
 PI

XX WPI; 2003-248177/24.

XX New thermostable Aviii peptide from Acidothermus cellulolyticus, useful
 PT for degradation of cellulose or in generating anti-Aviii antibodies for
 PT purifying recombinant Aviii polypeptides from genetically engineered
 PT host cells.

XX Claim 6; Page 8; 44pp; English.

XX The present sequence is derived from a thermostable avicelase, designated
 CC Aviii. Aviii is a member of the glycoside hydrolase family of enzymes,
 CC and is a cellulase. Aviii is useful in the conversion of biomass to
 CC biofuels and biofuel additives. It may be useful in the production of
 CC detergents, pulp and paper processing, food and feed processing and in
 CC textile processes. The thermostable Aviii peptide is useful in the
 CC degradation of cellulose, and in generating specific anti-Aviii
 CC antibodies that are useful in purifying recombinant Aviii polypeptides
 CC from genetically engineered host cells, in detecting Aviii polypeptide
 CC expression, as well as a reagent tool for characterizing the molecular
 CC actions of the polypeptide. The Aviii polynucleotide is useful as a
 CC source of probes or primers in various diagnostic assays

XX Sequence 89 AA;

Query Match 98.5%; Score 462; DB 6; Length 89;
 Best Local Similarity 97.7%; Pred. NO. 3.5e-45;
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

Db 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPTPTADTYLQ 88

Db 61 WAAIGCGNIRASFGSVNPTPTADTYLQ 88

RESULT 14

AD052316

ID AD052316 standard; protein; 89 AA.

XX AD052316;

XX 15-JUL-2004 (first entry)

XX A. cellulolyticus Aviii CBD_III partial protein #1.

XX Thermostable cellulase; avicelase III; Aviii; cellulose reduction;
 KW biomass degradation; ethanol formation; industrial chemical;
 KW fabric treatment; carbohydrate binding domain type III; CBD III.

XX

OS Acidothermus cellulolyticus.

XX Key Location/Qualifiers
 XX Misc-difference 89 /note= "A string of unknown amino acids (50 or more)"

XX US2004038334-A1.

XX 26-FEB-2004.

XX 28-JUL-2001; 2001US-00917376.

XX 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

XX (ADNEY/) ADNEY W S.

XX (VINZ/) VINZANT T B.

XX (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;
 PI

XX WPI; 2004-203224/19.

XX Novel thermostable Aviii polypeptide of glycoside hydrolase family and
 PT isolated from Acidothermus cellulolyticus, useful for degrading treated
 PT biomass into simpler forms of carbohydrate.

XX Claim 28; SEQ ID NO 4; 19pp; English.

XX The invention relates to a thermostable cellulase enzyme, avicelase III
 CC (Aviii) and its nucleic acid sequence. Aviii is useful for reducing
 CC cellulose in a starting material. A thermostable Aviii peptide is useful
 CC for degrading treated biomass into simpler forms of carbohydrate, which
 CC is used in the formation of ethanol or other industrial chemicals. It is
 CC also useful for treating fabrics to remove cellulose-containing stains.
 CC The present sequence is Acidothermus cellulolyticus Aviii carbohydrate
 CC binding domain type III (CBD_III) partial protein.

XX Sequence 89 AA;

Query Match 98.5%; Score 462; DB 8; Length 89;
 Best Local Similarity 97.7%; Pred. NO. 3.5e-45;
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

Db 1 VSGGVKQVYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

QY 61 WAAMCGNIRASFGSVNPTPTADTYLQ 88

Db 61 WAAIGCGNIRASFGSVNPTPTADTYLQ 88

RESULT 15

ABP73015

ID ABP73015 standard; protein; 957 AA.

XX AC ABP73015;

XX 03-JUN-2003 (first entry)

XX Amino acid sequence of the avicelase Aviii.

XX Avicelase; Aviii; glycoside hydrolase; enzyme; cellulase; biofuel;
 KW detergent; pulp processing; paper processing; feed processing; textile;
 KW cellulose.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

XX Misc-difference 957

XX /note= "unspecified residue encoded by N"

XX

PN WO2003012090-A2.
XX
PD 13-FEB-2003.
XX
XX
PF 28-JUL-2001; 2001WO-US023818.
XX
XX
PR 28-JUL-2001; 2001WO-US023818.
XX
XX
PA (MIDE) MIDWEST RES INST.
XX
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
XX
DR WPI; 2003-248177/24.
XX
XX
DR N-PSDB; ABZ77632.
XX
XX
PT New thermostable AvIII peptide from Acidothermus cellulolyticus, useful
PT for degradation of cellulose or in generating anti-AvIII antibodies for
PT purifying recombinant AvIII polypeptides from genetically engineered
PT host cells.
XX
XX
PS Claim 2; Page 20; 44pp; English.
XX
XX
CC The present sequence represents a thermostable avicelase polypeptide,
CC designated AvIII. AvIII is a member of the glycoside hydrolase family
CC of enzymes, and is a cellulase. AvIII is useful in the conversion of
CC biomass to biofuels and biofuel additives. It may be useful in the
CC production of detergent, pulp and paper processing, food and feed
CC processing and in textile processes. The thermostable AvIII peptide is
CC useful in the degradation of cellulose, and in generating specific anti-
CC AvIII antibodies that are useful in purifying recombinant AvIII
CC polypeptides from genetically engineered host cells, in detecting AvIII
CC polypeptide expression, as well as a reagent tool for characterizing the
CC molecular actions of the polypeptide. The AvIII polynucleotide is useful
CC as a source of probes or primers in various diagnostic assays
XX
SQ Sequence 957 AA;

Query Match 98.5%; Score 462; DB 6; Length 957;
Best Local Similarity 97.7%; Pred. No. 6.6e-44;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSTLVYNC 60
869 VSGGVKQYKNDAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSTLVYNC 928
Db
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956
Db

Search completed: October 5, 2005, 08:03:11
Job time : 16.8581 secs

Best Local Similarity 43.5%; Pred. NO. 1.4e-11;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSILVYNCDDAA 63
DB 1113 GLKLYLYKNNSTASTSRPFWKINGGSSVDLSTVRKIRYWTYVDGDKPOSAY-CDWAQ 1171
QY 64 MGCNGIRASFGSVNPAFTADTYLQ 88
DB 1172 IGASNVTFNFVKLTSVGVGADYYLE 1196

RESULT 3
G69593
cellulase (EC 3.2.1.4) bglC precursor - Bacillus subtilis
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: G69593; A26114; I40353; S24239; S49103; I39803
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruchi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
lech, J.; Harwood, C.R.; Hentut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y. M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleith, S.; Schoeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G69593
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <KUN>
A:Cross-references: UNIPROT:P10475; UNIPROT:Q03882; GB:Z99113; GB:AL009126; NID:G2634090
R:MacKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moranelli, F.; Sel
Nucleic Acids Res. 14, 9159-9170, 1986
A>Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
A:Reference number: A26114; MUID:87066783; PMID:3024130
A:Accession: A26114
A:Molecule type: DNA
A:Residues: 10-508 <MAC>
A:Cross-references: GB:X04689; NID:G39823; PIDN:CAA28392.1; PID:G39824
A:Experimental source: strain PAP115
A>Note: part of this sequence, including the amino end of the mature form, was confirmed
R:Lindahl, V.; Aa, K.; Tronemo, A.
Antonie Van Leeuwenhoek 66, 327-332, 1994
A>Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis C
A:Reference number: I40353; MUID:95225656; PMID:7710280
A:Accession: I40353
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN1>
A:Cross-references: EMBL:X67044; NID:G39776; PIDN:CAA47429.1; PID:G39777
R:Lindahl, V.; Aa, K.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24239
A:Accession: S24239
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN2>
A:Cross-references: EMBL:X67044; NID:G39776; PIDN:CAA47429.1; PID:G39777
A:Experimental source: strain CK-2
R:Wolf, M.; Geczi, A.; Borris, R.
submitted to the EMBL Data Library, December 1993
A>Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons
A:Reference number: S49103
A:Accession: S49103

A:Molecule type: DNA
A:Residues: 10-508 <WOL>
A:Cross-references: EMBL:Z29076; NID:G509266; PIDN:CAA2317.1; PID:G509267
R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.
Korean J. Microbiol. 24, 236-242, 1986
A>Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis
A:Reference number: I39803
A:Accession: I39803
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'P', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ', 36,
A:Cross-references: GB:M38634; NID:G142657; PIDN:AAA22300.1; PID:G142658
A:Experimental source: strain ATCC 6633
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
C:Genetics:
A:Gene: bglC
C:Function:
A>Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-38/Domain: signal sequence #status predicted <SIG>
F:39-508/Product: cellulase #status predicted <MAT>

Query Match 40.5%; Score 190; DB 2; Length 508;
Best Local Similarity 40.7%; Pred. No. 4.2e-12;
Matches 35; Conservative 21; Mismatches 26; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSILVYNCDDAA 62
DB 362 GVSQVYKAGDGRVNSNQIRPQLHKNGNTVDLKDVTARYWYKAKNGQN---FDCDYA 418
QY 63 AMGCNGIRASFGSVNPAFTADTYLQ 88
DB 419 QIGCGNVTHKFTLHKPKQGADTYLE 444

RESULT 4
A26874
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 09-Jul-2004
C:Accession: A26874; B26874
R:Robson, L.M.; Chambliss, G.H.
J. Bacteriol. 169, 2017-2025, 1987
A>Title: Endo-beta-1-4-glucanase gene of Bacillus subtilis DLG.
A:Reference number: A26874; MUID:87194581; PMID:3106328
A:Accession: A26874
A:Molecule type: DNA
A:Residues: 1-508 <ROB1>
A:Cross-references: UNIPROT:P07983; GB:M16185; NID:G143007; PIDN:AAA22496.1; PID:G143007
A:Experimental source: strain DLG
A:Accession: B26874
A:Molecule type: protein
A:Residues: 39-53 <ROB2>
A:Experimental source: strain DLG
A>Note: the authors believe Met-1 and Met-2 may be alternate initiators
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
C:Function:
A>Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; poly
F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 40.5%; Score 190; DB 2; Length 508;
Best Local Similarity 43.0%; Pred. No. 4.2e-12;
Matches 37; Conservative 17; Mismatches 28; Indels 4; Gaps 2;

QY 4 GVKVQYKNDSPAGDNQIKPGLQVNTGSSVDLSTVTYVWFTRDGGSSILVYNCDDAA 62
DB 362 GVSQVYKAGDGRVNSNQIRPQLHKNGNTVDLKDVTARYWYKAKNGQN---FDCDYA 418
QY 63 AMGCNGIRASFGSVNPAFTADTYLQ 88


```
Db 419 QMGCNLTHTKFTVTLHKPKQGADTYLE 444
|||||:|:::|||||:
2 SGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNCW 61
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
364 SQGIKVLVYANKETNSTNTIRPWLKVNSSGSSSIDLSRVTIRYWTVDGERAQSAIS-DW 422
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
62 AAMGCGNIRASFSGSVNPATPTADTYLQ 88
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
423 AQIGASNVTFKFKVLSSSVSGADYYLE 449
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 5
JN0111
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Bacillus subtilis
C/Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C/Accession: JN0111
R/Parc, S.H.; Kim, H.K.; Pack, M.Y.
Agricult. Biol. Chem. 55, 441-448, 1991
A/Title: Characterization and structure of the cellulase gene of Bacillus subtilis BSE61
A/Reference number: JN0111; MUID:91299280; PMID:1368694
A/Accession: JN0111
A/Molecule type: DNA
A/Residues: 1-499 <PAR>
A/Cross-references: UNIPROT:P23549; GB:D01057; NID:Q216387; PIDN:BAA00859.1; PID:di00132
A/Note: the authors translated the codon ATA for residue 102 as Tyr
C/Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A/Pathway: cellulose degradation
C/Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-29/Domain: signal sequence #status predicted <SIG>

Query Match 40.1%; Score 188; DB 2; Length 499;
Best Local Similarity 40.0%; Pred. No. 6.6e-12;
Matches 34; Conservative 18; Mismatches 31; Indels 2; Gaps 1;

QY 4 GVKVQYKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNCWAA 63
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
353 GISVQVRAGDGMNSNQIRPOLQIKNGNTVDLKDVTARYWYNAKNGQNV--DCDYAQ 410
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 64 MCGGNIRASFSGSVNPATPTADTYLQ 88
|||||:|:::|||||:
411 LGCGNVYTFKFTVTLHKPKQGADTYLE 435
|||||:|:::|||||:

RESULT 6
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N/Alternate names: beta-mannanase
C/Species: Caldocellum saccharolyticum
C/Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C/Accession: A48954; B43745
R/Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A/Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A/Reference number: A48954; MUID:93119139; PMID:1476429
A/Accession: A48954
A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1-1331 <GIB>
A/Cross-references: UNIPROT:P22533; GB:L01257; NID:G144230; PIDN:AAA71887.1; PID:G144291
A/Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBIPI:121577)
R/Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A/Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A/Reference number: A43745; MUID:91247819; PMID:2039230
A/Accession: B43745
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-337; PPRQHQHQ <LUE>
A/Cross-references: EMBL:M36063; NID:G144292; PIDN:AAA72861.1; PID:G144294
A/Note: the authors translated the codon CAC for residue 262 as Glu
A/Note: this sequence has been revised in reference A48954
C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 39.8%; Score 186.5; DB 2; Length 1331;
Best Local Similarity 42.5%; Pred. No. 2.7e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;
```

```
QY 2 SGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNCW 61
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
364 SQGIKVLVYANKETNSTNTIRPWLKVNSSGSSSIDLSRVTIRYWTVDGERAQSAIS-DW 422
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
62 AAMGCGNIRASFSGSVNPATPTADTYLQ 88
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
423 AQIGASNVTFKFKVLSSSVSGADYYLE 449
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 7
T17120
cellulase (EC 3.2.1.4) precursor, thermoactive - Caldocellum saccharolyticum
C/Species: Caldocellum saccharolyticum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T17120; A43745
R/Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A/Title: Cella, another gene coding for a multidomain cellulase from the extreme thermop
A/Reference number: Z18698; MUID:95336703; PMID:7612247
A/Accession: T17120
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1742 <TEO>
A/Cross-references: UNIPROT:P22534; EMBL:L32742; NID:G537499; PID:G537500; PIDN:AAA9108
R/Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A/Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene codin
A/Reference number: A43745; MUID:91247819; PMID:2039230
A/Accession: A43745
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1516-1544, A', 1546-1742 <LUE>
A/Cross-references: EMBL:M36063; NID:G144292; PIDN:AAA72860.1; PID:G144293
C/Genetics:
A/Gene: celsA
C/Keywords: glycosidase; hydrolase

Query Match 39.8%; Score 186.5; DB 2; Length 1742;
Best Local Similarity 42.5%; Pred. No. 3.6e-11;
Matches 37; Conservative 18; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGGVKQVQKNNDSAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSTLVYNCW 61
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
704 SQGIKVLVYANKETNSTNTIRPWLKVNSSGSSSIDLSRVTIRYWTVDGERAQSAIS-DW 762
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
62 AAMGCGNIRASFSGSVNPATPTADTYLQ 88
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
763 AQIGASNVTFKFKVLSSSVSGADYYLE 789
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 8
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IFO3034)
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Bacillus subtilis
C/Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 09-Jul-2004
C/Accession: A27198
R/Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A/Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A/Reference number: A27198; MUID:87190397; PMID:3106035
A/Accession: A27198
A/Molecule type: DNA
A/Residues: 1-499 <NAK>
A/Cross-references: UNIPROT:Q45532; GB:M28332; NID:G142670; PIDN:AAA22307.1; PID:G14267
A/Experimental source: strain IFO3034
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A/Pathway: cellulose degradation
C/Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <SIG>
```


Db 411 LGCNVSHVTTLHKPKQGAADTYLE 435

RESULT 13

B41897
cellulase (EC 3.2.1.4) - Bacillus lautus
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus lautus
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 09-Jul-2004
C:Accession: B41897; S27499
R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A:Title: ccla from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4
A:Reference number: A41897; MUID:92276330; PMID:1592807
A:Accession: B41897
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-700 <HAN>
A:Cross-references: UNIPROT:P29719; EMBL:W76588; NID:G142661; PIDN:AAA22303.1; PID:G14266
A:Experimental source: PL236
A:Note: sequence extracted from NCBI backbone (NCBIP:104605)
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.7%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 4.4e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSSTLVYNC 59
Db 549 VNSDLVQYKQDRNATDNQIKFNIQKGTSPVLSLTRYFTKD-SSAANGWI 607

QY 60 DNAAMGCNIRASFGSVNPATPTADTYLQ 88
Db 608 DWAKLGGNSIQISFGNHGA--DSDIYAE 634

RESULT 14

A47704
endoglucanase I (EC 3.2.1.-) Celli - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A47704
R:Hazlewood, G.P.; Davidson, K.; Laurie, J.I.; Huskisson, N.S.; Gilbert, H.J.
J. Gen. Microbiol. 139, 307-316, 1993
A:Title: Gene sequence and properties of Celli, a family E endoglucanase from Clostridium
A:Reference number: A47704; MUID:93171873; PMID:8436949
A:Accession: A47704
A:Molecule type: DNA; protein
A:Status: preliminary
A:Residues: 1-879 <HAZ>
A:Cross-references: UNIPROT:Q02934; GB:I04735; NID:G144807; PIDN:AAA20892.1; PID:G144808
A:Note: sequence extracted from NCBI backbone (NCBIN:125637, NCBIP:125638)
C:Keywords: glycosidase; hydrolase

Query Match 35.3%; Score 165.5; DB 2; Length 879;
Best Local Similarity 33.0%; Pred. No. 2.7e-09;
Matches 29; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSSTLVYNC 60
Db 736 IKGEVLQYANGAGTNSINPRFKIINNGTKAINLSDVKIRYYTYKGGASQNF-CD 794

QY 61 WAAMGCNIRASFGSVNPATPTADTYLQ 88
Db 795 WSSAGNSNVGTGNFNLSPKEGADTCLE 822

RESULT 15

S36859
cipA protein - Clostridium thermocellum

N:Alternate names: probable cellulosome protein large chain SL
C:Species: Clostridium thermocellum
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: S36859; S33527; S25767; S28659; T18261
R:Gerrigross, U.T.; Demain, A.L.
submitted to the EMBL Data Library, January 1993
A:Reference number: S36859
A:Accession: S36859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1854 <GER>
A:Cross-references: UNIPROT:Q06851; EMBL:L08665
R:Gerrigross, U.T.; Romaniec, M.P.M.; Kobayashi, T.; Huskisson, N.S.; Demain, A.L.
Mol. Microbiol. 8, 325-334, 1993
A:Title: Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal
A:Reference number: S33527; MUID:93302508; PMID:8316083
A:Accession: S33527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1615, 1617-1854 <GE2>
A:Cross-references: EMBL:L08665
R:Fujino, T.; Beguin, P.; Aubert, J.P.
FEMS Microbiol. Lett. 94, 165-170, 1992
A:Title: Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that
A:Reference number: S25767
A:Accession: S25767
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1854 <FUJ>
A:Cross-references: EMBL:X67406
R:Bequin, P.
submitted to the EMBL Data Library, August 1992
A:Reference number: S28659
A:Accession: S28659
A:Molecule type: DNA
A:Residues: 1216-1232, 'Y', 1234-1241, 'T', 1243-1319, 'R', 1321-1615, 1617-1768, 'R', 1770-1854
A:Cross-references: EMBL:X67406
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the celluloso
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18261
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1821-1854 <FU2>
A:Cross-references: EMBL:X67506; NID:G296879; PID:G296880; PIDN:CAA47840.1
C:Genetics:
A:Gene: cipA

Query Match 35.2%; Score 165; DB 2; Length 1854;
Best Local Similarity 36.1%; Pred. No. 6.7e-09;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKVOYKND-SAPGDNQIKPGLQVNTGSSVDLSTVTYRWFTRDGGSSTLVYNC 60
Db 365 VSGNLKVEFVNSPSTTNSINPQFKVTNTGSSAIDLSKLTLYRYTYVDGQKQDTFW-CD 423

QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQ 88
Db 424 HAAIIGNSGYNGITSNVKGTFVKMSSSTNNADTYLE 460

Search completed: October 5, 2005, 07:57:10
Job time : 3.44184 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:44:57 ; Search time 10.9413 Seconds
(without alignments)
4118.611 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKQYKKNDSAPGDNQ.....IRASFGSVNPAFTADTYLQ 88

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	257	54.8	741	Q82QF2	Q82qf2 streptomyc
2	228.5	48.7	616	Q7X2N2	Q7x2n2 thermomonos
3	225.5	48.1	170	Q9RFK6	Q9rfk6 caldibacill
4	216.5	46.2	930	Q9RFK5	Q9rfk5 caldibacill
5	214.5	45.7	921	Q918L8	Q918l8 caldibacill
6	209.5	44.7	1091	Q8XKF7	Q8xfk7 paenibacill
7	204.5	43.6	997	Q9Z4I1	Q9z4i1 bacillus sp
8	194.5	41.5	1751	Q9AQG4	Q9aqg4 caldicellul
9	192.5	41.0	1000	Q24820	Q24820 thermophil
10	192.5	41.0	1770	Q9X3P5	Q9x3p5 caldicellul
11	191.5	40.8	261	Q9AQG7	Q9aqg7 caldicellul
12	191.5	40.8	1426	Q9X3P6	Q9x3p6 caldicellul
13	191	40.7	1711	P96311	P96311 anaerocellu
14	190.5	40.6	996	Q9AQH0	Q9aqh0 caldicellul
15	190.5	40.6	1779	Q52374	Q52374 caldicellul
16	190	40.5	499	GUN1_BACSU	P07983 bacillus su
17	190	40.5	499	GUN2_BACSU	P10475 bacillus su
18	190	40.5	499	Q93TJ6	Q93tj6 bacillus su
19	190	40.5	508	Q93LD0	Q93ld0 bacillus su
20	190	40.5	508	Q70K05	Q70k05 bacillus am
21	188	40.1	499	GUN3_BACSU	P23549 bacillus su
22	187	39.9	499	Q52731	Q52731 bacillus sp
23	187	39.9	499	Q6Y4Q1	Q6y4q1 bacillus su
24	186.5	39.8	1331	MANE_CALSA	P22533 caldocellum
25	186.5	39.8	1742	GUNA_CALSA	P22534 caldocellum
26	186	39.7	499	Q45532	Q45532 bacillus su
27	186	39.7	501	Q83012	Q83012 bacillus sp
28	182.5	38.9	1039	GUNB_CALSA	P10474 c endogluc
29	180.5	38.5	145	YCEA_PAEIA	P29718 paenibacill
30	180	38.4	486	Q45430	Q45430 bacillus sp
31	177	37.7	499	Q8RPQ6	Q8rpq6 bacillus am

RESULT 1

ID	Q82QF2	PRELIMINARY;	PRT;	741 AA.
AC	Q82QF2;			
DT	01-JUN-2003 (Tremblrel. 24, Created)			
DT	01-JUN-2003 (Tremblrel. 24, Last sequence update)			
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)			
DE	Putative cellulose 1,4-beta-cellobiosidase.			
GN	Names=guxAI; OrderedLocusNames=SAV557;			
OS	Streptomyces avermitilis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycineae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=33903;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=MA-4680;			
RX	MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;			
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,			
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonce T.,			
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;			
RT	"Genome sequence of an industrial microorganism Streptomyces			
RT	avermitilis; deducing the ability of producing secondary			
RT	metabolites."			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).			
RN	[2]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=MA-4680;			
RX	MEDLINE=22608306; PubMed=12692562;			
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,			
RA	Sakaki Y., Hattori M., Omura S.;			
RT	"Complete genome sequence and comparative analysis of the industrial			
RT	microorganism Streptomyces avermitilis."			
RL	Nat. Biotechnol. 21:526-531(2003).			
CC	-I- SIMILARITY: Contains 1 fibronectin type III domain.			
EMBL	AP005023; BAC68267.1; ..			
HSSP	P20533; 1K85			
GO	GO:0004553; P:carbohydrase activity, hydrolyzing O-glycosyl . . . ; IEA.			
GO	GO:0005975; P:carbohydrate metabolism; IEA.			
InterPro	IPR001956; CBD_3.			
InterPro	IPR008965; Cellul bind.			
InterPro	IPR003961; FN_III-like			
InterPro	IPR008957; FN_III-like			
InterPro	IPR001524; Glyco_hydro_6.			
InterPro	IPR011253; Glyco_hydro_6-lk.			
Pfam	PF00942; CBM_3; 1.			
Pfam	PF00041; fn3; 1.			
Pfam	PF01341; Glyco_hydro_6; 1.			
PRINTS	PR00733; GLHYDRLAS6.			
ProDom	PD001947; CBD_3; 1.			
ProDom	PD003733; Glyco_hydro_6; 1.			
SMART	SM00060; FN3; 1.			
PROSITE	PS00853; FN3; 1.			
PROSITE	PS00655; GLYCOSYL_HYDROL_F6_1; 1.			
KW	Complete proteome.			


```
SQ SEQUENCE 741 AA; 77396 MW; 79404B40B2B47AF CRC64;
Query Match 54.8%; Score 257; DB 2; Length 741;
Best Local Similarity 56.3%; Pred. No. 1.5e-19;
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

Qy 2 SGGVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSSTLVYNCWD 61
Db 591 SGGKLVLYKNDSSATDNRPLGLRIVNTGSSLDLSKTVARYFTYFTRDGGSPVTNAWCY 650

Qy 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 651 AAVGCSNVSLKVPLITTPVPGADAYLE 677

RESULT 2
Q7X2N2 PRELIMINARY; PRT; 616 AA.
AC Q7X2N2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Endoglucanase.
GN Name=Cel5B;
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=TW51;
RA Posta K., Baki E., Kukolva J., Hornok L.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY298814; AAP56348.1; -.
DR HSP; Q45996; IG43.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF00150; Cellulase; 1.
DR ProDom; PD001947; CBD 3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN 1.
SQ SEQUENCE 616 AA; 67701 MW; 24FFC1E1A3F5639 CRC64;

Query Match 48.7%; Score 228.5; DB 2; Length 616;
Best Local Similarity 52.3%; Pred. No. 1.7e-16;
Matches 45; Conservative 16; Mismatches 22; Indels 3; Gaps 2;

Qy 2 SGGVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSSTLVYNCWD 61
Db 471 TGALEVYRYNLSAADDSQIAPGLRLVNTGSSVTDLADVEIHYFTNEPG-GTLQFTCDW 529

Qy 62 AAMGCGNIRASFGSVNPATPTADTYL 87
Db 530 AQVGCANVNASTSL--SAPGADTSL 553

RESULT 3
Q9RFK6 PRELIMINARY; PRT; 170 AA.
AC Q9RFK6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Caldicabacillus cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicabacillus.
OX NCBI_TaxID=74586;
RN [1];
RP SEQUENCE FROM N.A.
RC MEDLINE=20120520; PubMed=10653733; DOI=10.1128/AEM.66.2.664-670.2000;
RX
```

```
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicabacillus cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22273.1; -.
DR HSP; Q06851; INBC.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR Pfam; PF00942; CBM 3; 1.
DR ProDom; PD001947; CBD 3; 1.
DR KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F44E3A0B4 CRC64;

Query Match 48.1%; Score 225.5; DB 2; Length 170;
Best Local Similarity 47.7%; Pred. No. 8.3e-17;
Matches 41; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

Qy 3 GGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSSTLVYNCWD 62
Db 20 GSVVQYRAADTNAGDNQKPHFRIVNRGTSVPLSELTIYTYTD-GDKPQVNCDA 78

Qy 63 AAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 79 QVGCNSVRGSFVKLSGTGTGADYIE 104

RESULT 4
Q9RFK5 PRELIMINARY; PRT; 930 AA.
AC Q9RFK5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Multidomain beta-1,4-mannanase precursor.
GN Name=manA;
OS Caldicabacillus cellulovorans.
OC Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Caldicabacillus.
OX NCBI_TaxID=74586;
RN [1];
RP SEQUENCE FROM N.A.
RC MEDLINE=20120520; PubMed=10653733; DOI=10.1128/AEM.66.2.664-670.2000;
RX Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
RT Caldicabacillus cellulovorans and action of the recombinant enzyme on
RT kraft pulp.";
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22274.1; -.
DR HSP; Q9ZF13; 3MAN.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR004302; Chitin_binding_3.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM 3; 2.
DR Pfam; PF00150; Cellulase; 1.
DR Pfam; PF03067; Chitin_bind_3; 1.
DR ProDom; PD001947; CBD 3; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN 1.
KW Signal.
FT CHAIN 1 33 Potential.
FT SIGNAL 34 930 multidomain beta-1,4-mannanase.
SQ SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;

Query Match 46.2%; Score 216.5; DB 2; Length 930;
Best Local Similarity 46.5%; Pred. No. 5.9e-15;
Matches 40; Conservative 17; Mismatches 28; Indels 1; Gaps 1;

Qy 3 GGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRVYFTRDGGSSSTLVYNCWD 62
```



```

Qy      63  AMCGCNIRASFGSVNPATPTADTYLQ 88
          |||::|||
Db      830 WGCNINLRGSLVKLATGRTGADYYLE 855

RESULT 6
Q8KKXF7                                PRELIMINARY;    PRT;   1091 AA.
AC      Q8KKF7;
DT      01-OCT-2002 (T=EMBLrel. 22, Created)
DT      01-OCT-2002 (T=EMBLrel. 22, Last sequence update)
DT      01-OCT-2002 (T=EMBLrel. 25, Last annotation update)
DE      Cellulose 1,4-beta-cellobiosidase precursor (EC 3.2.1.91).
GN      Name=ccl48C;
OS      Paenibacillus sp. BP-23.
```



```
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR01701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR InterPro; IPR002345; Lipocalin.
DR Pfam; PF00942; CBM_3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD01947; CBD_3; 2.
DR SMART; SMO0060; FN3; 2.
DR PROSITE; PS0853; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34 Potential.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 43.6%; Score 204.5; DB 2; Length 997;
Best Local Similarity 43.7%; Pred. No. 1.3e-13;
Matches 38; Conservative 21; Mismatches 27; Indels 1; Gaps 1;

QY 2 SGVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSLTLYNCDW 61
Db 846 TGTLEQVYRSGSGSSNAVTPQFNKNTGQAIDLSTVKIRYFTKD-GTEELSFWCY 904

QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 905 AQVGSANVQGMFVAVNPAGKTADTYVE 931

RESULT 8
Q9AQG4 PRELIMINARY; PRT; 1751 AA.
AC Q9AQG4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D.; Reeves R.A.; Farrington G.K.; Anderson P.; Williams D.P.;
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078042; AAK06394.1; -.
DR HSSP; P37700; 1GB7.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR01701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00942; CBM_3; 4.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 4.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 41.5%; Score 194.5; DB 2; Length 1751;
Best Local Similarity 44.7%; Pred. No. 3.3e-12;
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSLTLYNCDW 63
Db 4 GVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSLTLYNCDW 63

Db 678 GVKVLYKNNETSASTGSRPMPFKIVNGSSVDLSRVKIRYWTYVDGDKPSAV-CDWAQ 736
QY 64 MCGNIRASFGSVNPATPTADTYLQ 88
Db 737 IGASNVTFNFVKLSSGSGADYYLE 761

RESULT 9
Q24820 PRELIMINARY; PRT; 1000 AA.
AC Q24820;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K.; Machida Y.; Hattori K.; Iijima S.;
RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
hydrolases).
DR EMBL; AB008029; BAA22939.1; -.
DR HSSP; Q06851; 1NBC.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul. bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00331; Cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PD00134; GLHYDLASE10.
DR ProDom; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1000 AA; 113264 MW; B9F659A56A752C6B CRC64;

Query Match 41.0%; Score 192.5; DB 2; Length 1000;
Best Local Similarity 43.7%; Pred. No. 2.9e-12;
Matches 38; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 2 SGVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSVTYVYWFTRDGGSSLTLYNCDW 61
Db 372 SQGIKLYANKETNSTNTIREPLKVVNTGSSSIDLSRVITRYWTYVDGDKQASAVS-DW 430

QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 431 AQIGASNVTFNFVKLSSGSGADYYLE 457

RESULT 10
Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC Q9X3P5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE XynA.
GN Name=xynA;
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
```


[illegible]


```
Db 413 GLKLVKNNETSASTSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 471
Qy 64 MCGCNIRASFGSVNPTPTADTYLQ 88
Db 472 IGASNVTFNFVKLSGSGADYYLE 496

RESULT 13
P96311 ID P96311 PRELIMINARY; PRT; 1711 AA.
AC P96311;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 1,4-beta-glucanase (Fragment).
GN Name=ce1a;
OS Anaerococcus thermophilus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Anaerococcus group;
OC Anaerococcus.
OX NCBI_TaxID=31899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z-1320;
RA MEDLINE=98154434; PubMed=9493383;
RX Zverlov V., Mahr S., Riedel K., Bronnemeier K.;
RT "Properties and gene structure of a bifunctional cellulolytic enzyme
RT (Cela) from the extreme thermophile Anaerococcus thermophilus with
RT separate glycosyl hydrolase family 9 and 48 catalytic domains.";
RL Microbiology 144:457-465(1998).
DR EMBL; Z86105; CAB06786.1; -.
DR PIR; T31337; T31337.
DR HSP; P37700; 1087.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR00556; Glyco_hydro_48.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_hydro_9.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR00209; Pept_S8_S53.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02011; Glyco_hydro_48; 1.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR00844; GLHYDRLASE48.
DR ProDom; PD001947; CBD_3; 3.
DR ProDom; PD011903; Glyco_hydro_48; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
FT NON_TER
SQ SEQUENCE 1711 AA; 189979 MW; E3E987CEB9CD0C21 CRC64;

Query Match 40.78; Score 191; DB 2; Length 1711;
Best Local Similarity 43.88; Pred. No. 7.8e-12;
Matches 39; Conservative 18; Mismatches 30; Indels 2; Gaps 2;

Qy 1 VSGG-VKQVKNNDSPGDNQIKPGLQVNTGSSVDLSVTYVYWTRDGGSTLVNVC 59
Db 683 VAGQIKVLVANKETSTTIRLWLVKVVNTGSSVDLSRVYIYWTVDGDKAQAIS- 741

Qy 60 DWAMGCGNIRASFGSVNPTPTADTYLQ 88
Db 742 DWAGIGASNVTFKFKVLSGSSGADYYLE 770

RESULT 14
Q9AQH0 ID Q9AQH0 PRELIMINARY; PRT; 996 AA.
AC Q9AQH0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
```

```
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl hydrolase 5 (Fragment).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078038; AAK06388.1; -.
DR HSP; Q06851; INEC.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00942; CBM_3; 1.
DR ProDom; PD001947; CBD_3; 1.
DR Hydrolase. 996
KW NON_TER
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match 40.6%; Score 190.5; DB 2; Length 996;
Best Local Similarity 43.5%; Pred. No. 4.7e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

Qy 4 GVKQVKNNDSPGDNQIKPGLQVNTGSSVDLSVTYVYWTRDGGSTLVNCDWAA 63
Db 844 GLKLVKNNETSASAGSIRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 902

Qy 64 MCGCNIRASFGSVNPTPTADTYLQ 88
Db 903 IGASNVTFNFVKLSGSGADYYLE 927

RESULT 15
Q52374 ID Q52374 PRELIMINARY; PRT; 1779 AA.
AC Q52374;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN Name=xynC;
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1;
RX MEDLINE=99283888; PubMed=10356996;
RA Morris D.D., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RA "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. strain
RT Rt69B.1.";
RL Extremophiles 3:103-111(1999).
CC -!- SIMILARITY: Belongs to cellulase family F (family 10 of glycosyl
CC hydrolases).
DR EMBL; AF036924; AAB95326.1; -.
DR PIR; T31085; T31085.
DR HSP; Q06851; INEC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR006584; CBD_IV.
```



```
DR InterPro; IPR005084; CBM_6.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR006710; Glyco_hydro_43.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF02018; CBM_4_9; 2.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR Pfam; PF04616; Glyco_hydro_43; 1.
DR PRINTS; PR00134; GLHYDRLASE10.
DR ProDom; PD001947; CBD_3; 3.
DR SMART; SM00606; CBD_IV; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
KW Glycosidase; Hydrolase; Xylan degradation.
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;

Query Match 40.6%; Score 190.5; DB 2; Length 1779;
Best Local Similarity 43.5%; Pred. No. 9.3e-12;
Matches 37; Conservative 14; Mismatches 33; Indels 1; Gaps 1;

QY 4 GVKQYKNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYNCDWAA 63
Db 1113 GLKLYKNNETSASTGSRPWFKIVNGGSSVDLSRVKIRYWTVDGDKPQSAV-CDWAQ 1171

QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88
Db 1172 IGASNVTFNFVKLTSGVSGADYYLE 1196
```

Search completed: October 5, 2005, 07:56:09
Job time : 11.9413 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:49:28 ; Search time 3.381 Seconds
(without alignments)
1942.949 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQVKNDSAPGNDQ.....IRASFGVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgm2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgm2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgm2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgm2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgm2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgm2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195.5	41.7	616	3	US-09-136-574A-47
2	194.5	41.5	1751	3	US-09-136-574A-44
3	191.5	40.8	1426	3	US-09-136-574A-43
4	172	36.7	700	2	US-07-862-588B-2
5	165.5	35.3	551	2	US-09-033-537A-1
6	165	35.2	167	5	PCT-US95-13813-9
7	165	35.2	476	4	US-09-339-159B-4
8	165	35.2	493	3	US-09-198-986-10
9	165	35.2	493	3	US-09-198-985A-12
10	165	35.2	493	3	US-09-694-531-12
11	165	35.2	493	4	US-09-670-141-10
12	165	35.2	493	4	US-10-072-152-12
13	155.5	33.2	1352	4	US-09-784-554B-2
14	150.5	32.1	1350	4	US-09-784-554B-4
15	120.5	25.7	531	2	US-07-862-588B-7
16	114	24.3	162	1	US-08-048-164A-2
17	114	24.3	162	1	US-08-460-457-2
18	114	24.3	162	1	US-08-460-458-2
19	114	24.3	162	2	US-08-460-455-2
20	114	24.3	162	2	US-08-330-394A-2
21	114	24.3	162	3	US-09-006-636-7
22	114	24.3	163	3	US-09-006-632-7
23	114	24.3	163	3	US-09-325-274-7
24	114	24.3	163	3	US-09-277-716-22
25	113	24.1	382	3	US-09-609-161B-22
26	113	24.1	382	4	US-09-609-161B-22
27	112	23.9	154	2	US-08-330-394A-29

28	112	23.9	156	2	US-08-330-394A-22	Sequence 22, Appl
29	72.5	15.5	272	4	US-09-902-540-11056	Sequence 11056, A
30	65.5	14.0	1566	4	US-09-581-472B-2	Sequence 2, Appl1
31	65	13.9	474	4	US-09-248-796A-25524	Sequence 25524, A
32	64	13.6	143	4	US-09-301-593-26	Sequence 26, Appl
33	64	13.6	428	3	US-09-118-319-5	Sequence 5, Appl1
34	64	13.6	464	1	US-08-353-400-36	Sequence 36, Appl
35	64	13.6	472	4	US-09-301-593-30	Sequence 30, Appl
36	63.5	13.5	1785	3	US-09-341-587-3	Sequence 3, Appl1
37	63	13.4	453	4	US-09-301-593-18	Sequence 18, Appl
38	63	13.4	472	4	US-09-301-593-43	Sequence 43, Appl
39	63	13.4	718	4	US-09-328-352-4640	Sequence 4640, Ap
40	63	13.4	1133	4	US-09-902-540-112243	Sequence 12243, A
41	62.5	13.3	1581	3	US-09-110-517-2	Sequence 2, Appl1
42	62.5	13.3	288	3	US-09-423-439-38	Sequence 38, Appl
43	62.5	13.3	389	4	US-09-902-540-14347	Sequence 14347, A
44	62.5	13.3	445	1	US-08-353-400-33	Sequence 33, Appl
45	62.5	13.3	641	4	US-09-687-538B-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1

US-09-136-574A-47

; Sequence 47, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 47:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 616 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-09-136-574A-47

Query Match 41.7%; Score 195.5; DB 3; Length 616;
Best Local Similarity 43.2%; Pred. No. 3.5e-13;
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

Qy 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 60
Db 1 MGSQVLYKNNETSASTGSRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 59

Qy 61 WAAMGCGNIRASFGSVNPTADTYLQ 88
Db 60 WAQIGASNVTFNFVKLSGSGVADYYLE 87

RESULT 2

US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1751 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match 41.5%; Score 194.5; DB 3; Length 1751;
Best Local Similarity 44.7%; Pred. No. 1.6e-12;
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63
Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63

Db 678 GVKVLYKNNETSASTGSRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 736

Qy 64 MGCENIRASFGSVNPTADTYLQ 88
Db 737 IGASNVTFNFVKLSGSGVADYYLE 761

RESULT 3

US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6294366e
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43

Query Match 40.8%; Score 191.5; DB 3; Length 1426;
Best Local Similarity 43.5%; Pred. No. 2.7e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

Qy 4 GVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGDKPQSAV-CD 63
Db 413 GLKVLKNNETSASTGSRPFKIVNGSSVDLSRVKIRYWTVDGDKPQSAV-CD 471

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

Qy 64 MGCENIRASFGSVNPTADTYLQ 88

Db 472 IGASNVTFNFVKLSGSGVADYYLE 496

US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linna
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862.588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-2
Query Match 36.7%; Score 172; DB 2; Length 700;
Best Local Similarity 44.9%; Pred. No. 1.6e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;
QY 1 VSGGVKQYKND-SAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGSSTLVYNC 59
Db 549 VNSDLVQYKCDRNATDNQIKPHFNQKGTSPVLSLTLRYFTKD--SSAAMNGWI 607
QY 60 DWAMGCGNIRASFGSNVNPATPTADTYLQ 88
Db 608 DWAKLGSNIQISFGNHGA--DSDTYAE 634
RESULT 5
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sch lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.

STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/033,537A
FILING DATE: 02-MAR-1998
CLASSIFICATION: 008
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 0993/95
FILING DATE: 08-SEP-1995
APPLICATION NUMBER: PCT/DK96/00364
FILING DATE: 03-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4492.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-033-537A-1
Query Match 35.3%; Score 165.5; DB 2; Length 551;
Best Local Similarity 37.9%; Pred. No. 6e-10;
Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;
QY 2 SGGVKVQYKND-SAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGSSTLVYNC 61
Db 402 TGNLVVQYKVGDTSATDNQMKPSFNKNNGTTPVNLGSLKLYFTKD-GTADMSASFDW 460
QY 62 AAMGCGNIRASFGSNVNPATPTADTYLQ 88
Db 461 AQIGASNVSAAP--ANFTGNTDTYVE 485
RESULT 6
PCT-US95-13813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd.
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely
; APPLICANT: Wilchek, Meir
; APPLICANT: Lamed, Raphael
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; TITLE OF INVENTION: PROTEINS AND USE THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Broadway and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

RESULT 11


```

; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-10-072-152-12

Query Match 35.2%; Score 165; DB 4; Length 493;
Best Local Similarity 36.1%; Pred. No. 6e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCD 60
Db 331 VSGNLKVFYNSPDSPTTNSINPQPKVTNTGSSAIDLSKLTLYYYTVDGQDQTFW-CD 389

QY 61 WAAM-----GCGNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAITGNSYNGITNSVKGFVFKVSSSTNNADTYLE 426

RESULT 13
US-09-784-554B-2
; Sequence 2, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schmorri, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1352
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-2

Query Match 33.2%; Score 155.5; DB 4; Length 1352;
Best Local Similarity 31.8%; Pred. No. 2.3e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

QY 4 GVKVOYKXNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCDWAA 63
Db 1200 GLLQTRTADTKVNDNHLNPFQIVNKGTTSPINELKIRIYYTIDGREQ-TFNCVAT 1258

QY 64 MCGCNIRASFGSVNPATPTADTYLQ 88
Db 1259 LSCSKLNGKLVKMKAAATGADYYLE 1283

RESULT 14
US-09-784-554B-4
; Sequence 4, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schmorri, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1350
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-4

Query Match 32.1%; Score 150.5; DB 4; Length 1350;

```

```

Best Local Similarity 31.8%; Pred. No. 8.1e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

QY 4 GVKVOYKXNDSPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCDWAA 63
Db 1199 GVLQTRTADTKVNDNHLNPFQIVNKGTTSPINELKIRIYYTIDGREQ-TFNCVAT 1257

QY 64 MCGCNIRASFGSVNPATPTADTYLQ 88
Db 1258 LSCSKLNGKLVKMKAAATGADYYLE 1282

RESULT 15
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T.; Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-7

Query Match 25.7%; Score 120.5; DB 2; Length 531;
Best Local Similarity 32.9%; Pred. No. 5e-05;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

QY 2 SGGVKQVYKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGGSSTLVNCDW 61
Db 432 TGNLVVQYKVGDTSATDNQMKPSFNKKNNGTTPVNL-SGLKXXXXKD-GPADMSCSIDW 490

QY 62 AAMCGCNIRASFGSVNPATPTADTY 86
Db 491 AQIGRTNVLAF--ANFTGSNTDITY 513

Search completed: October 5, 2005, 08:04:32

```


us-09-917-376-5.ra1

Thu Oct 6 11:20:49 2005

Job time : 4.381 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 07:52:23 ; Search time 11.1761 Seconds
(without alignments)
3288.602 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKQYKNDSPAGDNOI.....IRASFGSYNPATPTADTYLQ 88

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues
Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10A_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	469	100.0	88	11	US-09-917-376-5
2	469	100.0	88	14	US-10-155-400-5
3	469	100.0	89	11	US-09-917-376-4
4	469	100.0	89	14	US-10-155-400-4
5	469	100.0	154	10	US-09-917-378-4
6	469	100.0	762	10	US-09-917-378-1
7	466	99.4	150	10	US-09-917-384-5
8	466	99.4	150	10	US-09-917-383-5
9	466	99.4	1043	10	US-09-917-384-6
10	466	99.4	1043	10	US-09-917-383-6
11	466	99.4	1228	10	US-09-917-384-1

12	466	99.4	1228	10	US-09-917-383-1	Sequence 1, Appli
13	462	98.5	957	11	US-09-917-376-1	Sequence 1, Appli
14	462	98.5	957	14	US-10-155-400-1	Sequence 1, Appli
15	257	54.8	741	14	US-10-156-761-8100	Sequence 8100, Ap
16	190	40.5	508	15	US-10-369-493-23151	Sequence 23151, A
17	188	40.1	477	16	US-10-466-208-12	Sequence 12, Appli
18	188	40.1	496	16	US-10-466-208-8	Sequence 8, Appli
19	188	40.1	677	16	US-10-433-577-35	Sequence 35, Appli
20	180	38.4	1621	14	US-10-185-990-10	Sequence 10, Appli
21	165	35.2	167	17	US-10-933-404-4	Sequence 4, Appli
22	165	35.2	476	15	US-10-372-054-4	Sequence 4, Appli
23	165	35.2	493	13	US-10-072-152-12	Sequence 12, Appli
24	165	35.2	493	15	US-10-655-433-12	Sequence 12, Appli
25	165	35.2	599	10	US-09-955-555A-29	Sequence 29, Appli
26	155.5	33.2	1352	10	US-09-784-554B-2	Sequence 2, Appli
27	155.5	33.2	1352	16	US-10-896-555-2	Sequence 2, Appli
28	150.5	32.1	1350	10	US-09-784-554B-4	Sequence 4, Appli
29	150.5	32.1	1350	16	US-10-896-555-4	Sequence 4, Appli
30	130.5	27.8	1483	15	US-10-282-122A-51483	Sequence 51483, A
31	114	24.3	256	14	US-10-261-446-6	Sequence 6, Appli
32	114	24.3	256	15	US-10-261-445B-6	Sequence 6, Appli
33	114	24.1	256	17	US-10-782-234-6	Sequence 6, Appli
34	113	24.1	382	10	US-09-808-898-22	Sequence 22, Appli
35	105.5	22.5	163	15	US-10-460-524-2	Sequence 2, Appli
36	78.5	16.7	1049	15	US-10-282-122A-49900	Sequence 49900, A
37	75.5	16.1	346	16	US-10-437-963-179234	Sequence 223, App
38	75.5	16.1	618	15	US-10-211-462-223	Sequence 4039, Ap
39	75.5	16.1	618	16	US-10-723-860-4039	Sequence 5713, Ap
40	75.5	16.1	618	18	US-10-756-149-5713	Sequence 53, Appli
41	75.5	16.1	986	9	US-09-747-835A-53	Sequence 101, App
42	75.5	16.1	986	14	US-10-120-604-101	Sequence 406, App
43	75.5	16.1	986	14	US-10-225-567A-406	Sequence 3, Appli
44	75.5	16.1	986	15	US-10-398-458-3	Sequence 53, Appli
45	75.5	16.1	986	15	US-10-312-312-53	

ALIGNMENTS

RESULT 1
US-09-917-376-5
; Sequence 5, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURES:
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-5

Query Match	100.0%;	Score 469;	DB 11;	Length 88;
Best Local Similarity	100.0%;	Pred. No. 3.6e-48;		
Matches	88;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	VSGGVKQYKNDSPAGDNOIKPGLQIVNTGSSVDLSTVTVYWFTRDGGSSILVYNC	60	
Db	1	VSGGVKQYKNDSPAGDNOIKPGLQIVNTGSSVDLSTVTVYWFTRDGGSSILVYNC	60	
Qy	61	WAAMGCNIRASFGSYNPATPTADTYLQ	88	

[illegible]

OTHER INFORMATION: Guxa
US-09-917-384-5
Query Match 99.4%; Score 466; DB 10; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.6e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTTRDGGSTLVYVNC 60
Db 1 VSGGLKVQYKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTTRDGGSTLVYVNC 60
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
RESULT 8
US-09-917-383-5
; Sequence 5, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-5
Query Match 99.4%; Score 466; DB 10; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.6e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTTRDGGSTLVYVNC 60
Db 1 VSGGLKVQYKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTTRDGGSTLVYVNC 60
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
RESULT 9
US-09-917-384-6
; Sequence 6, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6

Query Match 100.0%; Score 469; DB 10; Length 154;
Best Local Similarity 100.0%; Pred. No. 7e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTTRDGGSTLVYVNC 60
Db 1 VSGGVKQVQKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTTRDGGSTLVYVNC 60
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
RESULT 6
US-09-917-378-1
; Sequence 1, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
US-09-917-378-1
Query Match 100.0%; Score 469; DB 10; Length 762;
Best Local Similarity 100.0%; Pred. No. 4.9e-47;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VSGGVKQVQKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTTRDGGSTLVYVNC 60
Db 455 VSGGVKQVQKNNDSAPGDNQIKPGLQVNTGSSVDLSTVTYRYWFTTRDGGSTLVYVNC 514
QY 61 WAAMGCGNIRASFGSVNPATPTADTYLQ 88
Db 515 WAAMGCGNIRASFGSVNPATPTADTYLQ 542
RESULT 7
US-09-917-384-5
; Sequence 5, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of


```
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-6

Query Match          99.4%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.6e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 477 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 536
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564
   |||:|||||

RESULT 10
US-09-917-383-6
; Sequence 6, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-6

Query Match          99.4%; Score 466; DB 10; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.6e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 477 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 536
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564
   |||:|||||

RESULT 11
US-09-917-384-1
; Sequence 1, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
```

```
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-384-1

Query Match          99.4%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 2e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 584 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 643
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671
   |||:|||||

RESULT 12
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-1

Query Match          99.4%; Score 466; DB 10; Length 1228;
Best Local Similarity 98.9%; Pred. No. 2e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 60
   |||:|||||
Db 584 VSGGLKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGGSSTLVNCD 643
   |||:|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
   |||:|||||
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671
   |||:|||||

RESULT 13
US-09-917-376-1
; Sequence 1, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
```



```
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917.376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
US-09-917-376-1

Query Match      98.5%; Score 462; DB 11; Length 957;
Best Local Similarity 97.7%; Pred. No. 4.4e-46;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVYVNC 60
Db 869 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVYVNC 928

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 14
US-10-155-400-1
; Sequence 1, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (957)
; OTHER INFORMATION: Any amino acid
US-10-155-400-1

Query Match      98.5%; Score 462; DB 14; Length 957;
Best Local Similarity 97.7%; Pred. No. 4.4e-46;
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVYVNC 60
Db 869 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVYVNC 928

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 15
US-10-156-761-8100
; Sequence 8100, Application US/10156761
```

```
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8100
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8100

Query Match      54.8%; Score 257; DB 14; Length 741;
Best Local Similarity 56.3%; Pred. No. 1.2e-21;
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 2 SGGVKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVYVNC 61
Db 591 SGGKLVLYKNDSSATDINAIRPGLRIVNTGSGSLDLSKVTRYFFTRDGSPTVNAWCDY 650

QY 62 AAAMCGNIRASFGSVNPATPTADTYLQ 88
Db 651 AAAGCSNVSLKVVPVLTTPVPGADAYLE 677

Search completed: October 5, 2005, 08:08:32
Job time : 11.1761 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 07:09:01 ; Search time 1158.46 Seconds
(without alignments)
3680.789 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKQYKXNDAPGDQ.....IRASFGSVNPATPTADTVLQ 88

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09917376/runat_04102005.164328.26897/app.query.fasta.1.2524
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CNC 1 1.15287 @runat_04102005.164328.26897 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	469	100.0	2289	6 AX700050	AX700050 Sequence
2	466	99.4	3687	6 AX700036	AX700036 Sequence
3	462	98.5	2869	6 AX700058	AX700058 Sequence
4	436	93.0	3365	6 AX700025	AX700025 Sequence

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	257	54.8	299175	1	AP005023	AP005023 Streptomy
6	228.5	48.7	1957	1	AX298814	AX298814 Thermobif
7	225.5	48.1	4567	1	AF163837	AF163837 Caldicell
8	214.5	45.7	3237	1	AF200304	AF200304 Caldibaci
9	209.5	44.7	3509	1	PAE488933	AJ488933 Paenibaci
10	204.5	43.6	4161	1	BSP133614	AJ133614 Bacillus
11	195.5	41.7	2029	6	E35142	E35142 Truncated c
12	195.5	41.7	2029	6	E35143	E35143 Truncated c
13	194.5	41.5	6005	1	AF078038S5	AF078042 Caldicell
14	194.5	41.5	6005	1	AF078038S5	E35100 Truncated c
15	192.5	41.0	4743	1	AB008029	AB008029 Thermophi
16	192.5	41.0	11707	1	AF078737	AF078737 Caldicell
17	192.5	41.0	11707	6	E35099	E35099 Truncated c
18	191.5	40.8	787	1	AF078038S3	AF078040 Caldicell
19	191	40.7	5513	1	AT286105	Z86105 A.thermophi
20	190.5	40.6	3262	1	AF078038S1	AF078038 Caldicell
21	190.5	40.6	5437	1	CASR69XN2	AF036924 Caldicell
22	190	40.5	1523	1	AF355629	AF355629 Bacillus
23	190	40.5	1800	1	BSEGLSG	Z29076 Bacillus su
24	190	40.5	1920	1	BACGLUB	M16185 B.subtilis
25	190	40.5	2314	1	BS14GLUC	X67044 B.subtilis
26	190	40.5	2435	1	BSRGLUC2	X04689 Bacillus su
27	190	40.5	2589	1	AY044252	AY044252 Bacillus
28	190	40.5	26170	1	BC170DEGR	Z73234 B.subtilis
29	190	40.5	122883	1	RAM575417	AJ575417 Bacillus
30	190	40.5	122883	1	RAM575417	AJ576102 Bacillus
31	190	40.5	207829	1	BSUB0010	Z99113 Bacillus su
32	188	40.1	1928	1	BACCMCASE	D01057 B.subtilis
33	187	39.9	1593	1	AY183475	AY183475 Bacillus
34	187	39.9	2084	1	AF045482	AF045482 Bacillus
35	186.5	39.8	4977	1	CDQWANA	L01257 Caldocellum
36	186.5	39.8	5439	1	CDCELA	L32742 Caldocellum
37	186	39.7	2175	1	BACCELD	M28332 B.subtilis
38	186	39.7	2175	6	E05425	E05425 DNA sequenc
39	186	39.7	2587	1	AB016164	AB016164 Bacillus
40	182.5	38.9	2977	6	A28170	A28170 B.lautus st
41	182.5	38.9	4241	1	CSCBLB	X13602 Caldocellum
42	182.5	38.9	5284	1	CSU16308	U16308 Caldocellum
43	180.5	38.5	2831	1	BACCELA	M76588 Bacillus la
44	180	38.4	1553	1	BSU27084	U27084 Bacillus sp
45	177	37.7	2152	1	AF363635	AF363635 Bacillus

ALIGNMENTS

RESULT 1	AX700050	Sequence 2 from Patent WO03012110.	2289 bp	DNA	linear	PAT 03-APR-2003
LOCUS	AX700050	Sequence 2 from Patent WO03012110.				
DEFINITION	AX700050					
ACCESSION	AX700050					
VERSION	AX700050.1	GI:29536020				
KEYWORDS						
SOURCE						
ORGANISM						
		Acidothermus cellulolyticus				
		Acidothermus cellulolyticus				
		Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;				
		Frankineae; Acidothermaceae; Acidothermus.				
REFERENCE						
AUTHORS		Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.				
TITLE		Thermal tolerant mannanase from Acidothermus cellulolyticus				
JOURNAL		Patent: WO 03012110-A 2 13-FEB-2003;				
		Midwest Research Institute (US)				
FEATURES						
source		Location/Qualifiers				
		1..2289				
		/organism="Acidothermus cellulolyticus"				
		/mol_type="unassigned DNA"				
		/db_xref="taxon:28049"				

ORIGIN

Alignment Scores:			
Pred. No.:	3.59e-48	Length:	2289
Score:	469.00	Matches:	88
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

CDS	YGIASLSYLSVGVNDARYLAEQALHTILLQWRITETPANGRPASS" 2694..3098 /note="SAV449" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68159.1" /db_xref="GI:29604088" /translation="MWYHINRVTKATPQIEGVLSWRNQGRSNPAIKSFVVGRN HGGDYTCAGVAVVHLDGLFAVLTHPTTYQTDHLGLHVERLEIFDVSDDDDDPDLNAX IQELHRRNELNPQIAGMLADVPYTGSGVDD" 3127..4029 /note="SAV450" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68160.1" /db_xref="GI:29604099" /translation="MAKHVIIGAGSIGTNVARLLAEGGESVRIIVTRSGGPEHPLVDR VAADASPSRLTELSTSGAGVITYHCANPPSYTWERLLPPLQTAATIAAKANDAVIALT GSLVAYGPQGRMNEHTPMAATGHKGLRRRWEOALAGGIRTVVEVRGSDYIGKDN GIYSLIKSALQKGAWITGHLDMPHFTFNGDMAQALITLASERAWGRPMWVPSA PAVIRELARYAVNAGRPPIKLQIPRSVTRTAGLIVPIAREMAEMDYQMYAPFHM ATETTFGLTATDLDTAVRNEVG" complement(4130..4498) /gene="bldg" complement(4130..4498) /gene="bldg" /note="SAV451" /codon_start=1 /transl_table=11 /product="putative anti-anti-sigma factor" /protein_id="BAC68161.1" /db_xref="GI:29604090" /translation="MTDAFHIDVPHDDAALAVIALSGSFDIAAPVAPARALELIANG HPDLVADLSGYTFCDSGLGALVGIWRCAKDAGSLTLAAIPDLRLSLVSTGMDAFL PAYSSADAALAAARQGNRTTA" 4635..5156 /note="SAV452" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68162.1" /db_xref="GI:29604091" /translation="MFAVKESRVVVGIEIATPTGEALELVLHGESLRGVRSDDA TARPDGLFELPARGWHITVALVPSEEGEAGGLFHPGAEPGESGPARLCAWRE SAPRTITDEPVEVHTLRSLIEDLGAPHLRPAAPRTHGRRRCPLVHGVRPRLLAAV LVHGTAAGPCTPR" 5183..5665 /note="SAV453" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68163.1" /db_xref="GI:29604092" /translation="MVTAVRARAGSAWAQCRGARRSSQHSSEGVREWDKEFLRTV AERTGLSQEADLTWATLETLARLSPEARDLITELPEGLAEAVRGTDRIFRG HDSVRRVAETMLKEEADRGVRTVLAVLEAISEKEFNDLMSQLGTFDSQALESAG " 5810..6964 /note="SAV454" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68164.1" /db_xref="GI:29604093" /translation="WIMTADVPTSSGRSVRVALVADPGLPSEIARALAPRLPHLRR AVGARVRQVDTWAPLSVSEQADVSDIABAVRSHLDGAENDVAIFLTFDFRRARLYP ISVEVDTRAAALISLPALGVRLLRRRVQAVDVRBELVADWHEPTPRPEVIGRPPE FVPDTRPQEPQVRVPGRLGHTLTVAGMYANRPWRLFGLSRLAGHPVATATVLF INSTATNIALGAVQHAVITAGSAVALTWIIVDHLQWREGTHLPARHHPLYPNLV TLVTITALAVVLAALFATLVAVSFLILDASVLRKFIGOVGVGDFLFWFITAIM IGGAGTSLGDEKVDAAVGRQRQRQRLRQMPRPERGERGPDITDG"
CDS	complement(6986..7498) /note="SAV455" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68165.1" /db_xref="GI:29604094" /translation="MNPYAPASGIBQRFQWMLDEYLMWCSCDIIADLAERGLRLEQPLV /trARLEAGERVLA VGPQORVTWRALGDRYMHQNAVAFGSPAFVLGSMASWATSALG NSARERAANDAQPRCGARTPGLHVLVDLVHGTMLRWSMLHGTSSSTRPSSALRALL ELVDLHVIAL" 7548..8189 /note="SAV456" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68166.1" /db_xref="GI:29604095" /translation="MATLLRRVCHEGVRIIAVPAGFVLSRRPWTASVYRCVAVVGQVS EDRREVEGSRTPVASASRLRPCCRLLRSVRCPRRAVPSQAVRSRFPRTGTRFLGDLV VYQRHDLAHARVRHVIDDVHPLGRLGRGAARPVRIATHSLGGVIAVDMATANVPLMTES LVTPGSQATFFHQPVPGDLLCQMIATFRMRKHQEAETGALPETGTIFAHAGQGW"
CDS	complement(6986..7498) /note="SAV455" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68165.1" /db_xref="GI:29604094" /translation="MNPYAPASGIBQRFQWMLDEYLMWCSCDIIADLAERGLRLEQPLV /trARLEAGERVLA VGPQORVTWRALGDRYMHQNAVAFGSPAFVLGSMASWATSALG NSARERAANDAQPRCGARTPGLHVLVDLVHGTMLRWSMLHGTSSSTRPSSALRALL ELVDLHVIAL" 7548..8189 /note="SAV456" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC68166.1" /db_xref="GI:29604095" /translation="MATLLRRVCHEGVRIIAVPAGFVLSRRPWTASVYRCVAVVGQVS EDRREVEGSRTPVASASRLRPCCRLLRSVRCPRRAVPSQAVRSRFPRTGTRFLGDLV VYQRHDLAHARVRHVIDDVHPLGRLGRGAARPVRIATHSLGGVIAVDMATANVPLMTES LVTPGSQATFFHQPVPGDLLCQMIATFRMRKHQEAETGALPETGTIFAHAGQGW"
gene	Alignment Scores: Pred. No.: 4,31e-19 Length: 299175 Score: 257.00 Matches: 49 Percent Similarity: 71.26% Conservative: 13 Best Local Similarity: 56.32% Mismatches: 25 Query Match: 54.80% Indels: 0 DB: 1 Gaps: 0
CDS	US-09-917-376-5 (1-88) x AP005023 (1-299175) Qy 2 SerGlyGlyValIleValGlnTyrIlyshenAspSerAlaProGlyAspGlnIle 21 Db 105341 TCGGCGCGCTCAAGGTCTCTACAGAACCAACAGCTCTCGGCCACCCGACCGCATC 105400 Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41 Db 105401 CGGCCAGCGCTTCGATGTCACACCGGACGCGCTCCCTCGACTGTCCAAGGTACG 105460 Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61 Db 105461 GCCCTTACTACTTCAACCGGACAGCGCTCGCCACCGTGAACGCTGGTGGCGACTAC 105520 Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81 Db 105521 GCGGCGCTCGGTGTTCACACGCTCAGCGCTGAGGTCTACCCCTGACCAACGCCCGGTGCCG 105580 Qy 82 ThrAlaAspThrTyrLeuGln 88 Db 105581 GGAGCCGACGCTACCTCGAA 105601 RESULT 6 LOCUS AY298814 Thermobifida fusca strain TMS1 endoglucanase (cel5B) gene, complete 1957 bp DNA linear BCT 23-JUN-2003 DEFINITION Thermobifida fusca strain TMS1 endoglucanase (cel5B) gene, complete cds. ACCESSION AY298814 VERSION AY298814.1 GI:31745732 KEYWORDS SOURCE Thermobifida fusca ORGANISM Thermobifida fusca Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida. REFERENCE 1 (bases 1 to 1957) AUTHORS Posta,K., Beki,E., Kukolya,J. and Hornok,L. TITLE Phylogenetic relationships of Tf cel5B, a new endoglucanase encoding gene from thermobifida fusca JOURNAL Unpublished REFERENCE 2 (bases 1 to 1957) AUTHORS Posta,K., Beki,E., Kukolya,J. and Hornok,L. TITLE Direct Submission

JOURNAL Submitted (14-MAY-2003) Agricultural Biotechnology and Microbiology, Szent Istvan University Godollo, Pater Karoly str.1., Godollo 2103, Hungary

FEATURES Location/Qualifiers
source 1..1957
/organism="Thermobifida fusca"
/mol_type="genomic DNA"
/strain="TW51"
/db_xref="taxon:2021"

misc_binding 26..39
/note="putative"
/bound_moiety="CelR"

gene 107..1957
/gene="cel5B"

CDS 107..1957
/gene="cel5B"

/transl_table=1
/product="endoglucanase"

/protein_id="AAP56348.1"

/db_xref="GI:31745733"

/translation="MTPLTRLRAGAAAIATGASALIPLTSSPAAAGTADWLHTDGN
RIVDSAGNEVLTCANWFNTSERMEFGLWAANIDITSAERGINMVRPISTOL
LLEKNGCAGSGVNEVYNPLAGMNTLEVDYMLCEEYGLKVLMDVSAEADNSG
HYIPWYKGDITTFYTANERWTERYKNDTIIVADIKNPHGKANETPRKWDGST
DIDNFKVCTAGKRIALAINPMLICGIBIIPDQGDSSDGRDITFTWNGNLR
GADHPDVLGADQGVSPHYDGSFVEQFEFEGEWNRRQTLEDVRPNWLYIHEDD
IAPLLIGMGQFLGDGKWTALRSIIIDEKMHFTFALNPNSGDTGGLNLTWHT
WDEAKYFLKALMDQANGKFGVLDHVPGLGVSTGVSLNOYGGGGPQPPTEPT
EPTEPTEPTEPPEPPANPTGALVYVYNNLSAADDSQIAPGLRVNTGSSTVDLAD
VEIHYFTFEPGTLQFTCDWAQVCANVNASFTSLAPGADTSLVLTLSGLARGAS
TELQRIHTANWANFDESDDYSRGTNTDWEISELVITAYLGGTLVWGTTPPA"

misc_feature 236..1258
/gene="cel5B"

/note="Region: catalytic domain; GH5 family"

misc_feature 1532..1765
/gene="cel5B"

/note="Region: cellulose binding domain; CBM-3"

ORIGIN

Alignment Scores:
Pred. No.: 2,786-18 Length: 1957
Score: 228.50 Matches: 45
Percent Similarity: 70.93% Conservative: 16
Best Local Similarity: 52.33% Mismatches: 22
Query Match: 48.72% Indels: 3
DB: 1 Gaps: 2

US-09-917-376-5 (1-88) x AY298814 (1-1957)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21

Db 1517 ACGGGTGGCTCGAGTCTACTACCCCAACACAGCTCTTGGCGCCGACGACGATC 1576

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41

Db 1577 GCGCGGGGCTGGCTGGTCAACACCGGAGCAGCAGCTCGACCTGCTGACGCTGAA 1636

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61

Db 1637 ATCCACTACTACTTACCAACAGCGCCGCGGCGGCGGCGGCGGCGGCGGCGG 1693

QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81

Db 1694 GCTCAAGTGGGCTGGCGCCCAACGCTCAACGCGTCTCTTCCGCTCGCTG-----TCGCTCGC 1747

QY 82 ThrAlaAspThrTyrLeu 87

Db 1748 GCGCGCGACACCTCCCTG 1765

RESULT 7

AF163837

LOCUS 4567 bp DNA linear BCT 08-FEB-2000

DEFINITION

Caldbacillus cellulovorans multidomain beta-1,4-mannanase precursor (mana) gene, complete cds; and unknown genes.

ACCESSION AF163837

VERSION AF163837.1 GI:6651325

KEYWORDS Caldbacillus cellulovorans

ORGANISM Caldbacillus cellulovorans

REFERENCE 1 (bases 1 to 4567)

AUTHORS Sunna,A., Gibbs,M.D., Chin,C.W., Nelson,P.J. and Bergquist,P.L.

TITLE A gene encoding a novel multidomain beta-1,4-mannanase from

Caldbacillus cellulovorans and action of the recombinant enzyme on

kraft pulp

JOURNAL Appl. Environ. Microbiol. 66 (2), 664-670 (2000)

MEDLINE 20120520

PUBMED 10653733

REFERENCE 2 (bases 1 to 4567)

AUTHORS Sunna,A., Gibbs,M.D. and Bergquist,P.L.

TITLE Direct Submission

JOURNAL Submitted (29-JUN-1999) Biological Sciences, Macquarie University,

North Ryde, New South Wales 2109, Australia

FEATURES Location/Qualifiers

source 1..4567

/organism="Caldbacillus cellulovorans"

/mol_type="genomic DNA"

/db_xref="taxon:74586"

CDS <1..513

/note="orf1"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAF22273.1"

/db_xref="GI:6651326"

/translation="TPTPTPSAAPTPSPAGSLVQVRAADTNAGDNQLKPHFRI

VNRGTSSVPLSELTIYWTVDKDPQVFNCDWAQVCSNVRGSLFKLSTGRGADYY

IEITFTSGAGSLAAGASSGDIQVRINKNDWTNNEANDYSYDPTKTSFADNRVTLR

NGQLLWGVEP"

misc_feature 1..60

/note="Region: proline-threonine linker"

/evidence=not_experimental

misc_feature 61..510

/note="Region: cellulose-binding domain type IIIB"

/evidence=not_experimental

gene 620..3412

/gene="mana"

620..3412

/gene="mana"

/codon_start=1

/transl_table=11

/product="multidomain beta-1,4-mannanase precursor"

/protein_id="AAF22274.1"

/db_xref="GI:6651327"

/translation="MNRRLIARLSGLAMVLIAMLAYVPKPPEAAGHGMVFPATRT

YACYDGVKVGNGDLNMINPACLDALISGNQVFNWFGNLIAGGRHREIIPDGK

LCGPASFDGMNQARTDWTNTRLPQCATITVRNAPHPGTWLYVTRDGDPTPL

KNSDLPTFPFQINSDDGAGYSQVQPNKQGHILYIMIQWSDSPAFYN

CSDVYFGSGPIAYEFQDPREGGTMITPPSGTTPPTPTPTPTPTPTPTPTPTPT

TPSTPTPSASGTLRVEYRGDTSATDNQKPOLRIVNTGSQAVPLTELKVRVYTKN

STQAOYFCDMAQIGCSNIRAQPVSLQSPVSGADSYIELSTGSGIPAGNGTGIQNR

IFHTWMYNETDDWSYNGAQTTWGPSTRIITLVNGLVWGTGPGGSGSPPTPTPT

GANAVRILNSNGCRWTKIPASEVADIIISQARTIGYRAVLEVDHDTTIGEDAAACMT

TANVYIELKNVLQAGENVIVNGNEPYNNGYQNVMTDTRNAVQALRAGNNTIM

VAPNGQWDSFTMRDNPATIFNADPQNLVFSIHMYGVYDTAAEVQSYIESFNRLG

PLVIGFEGHSHSDPNEQALVOYAKYNIGLFWSGNGGSGVYELMDVMTFNANSPL

TANGWFRNTAIGTSTPT

KGPHRLVNRGTSSVPLSELTIYWTVDKDPQVFNCDWAQVCSNVRGSLFKLSTGR

TCGADYYIEITFTSGAGSLAAGASSGDIQVRINKNDWTNNEANDYSYDPTKTSFADNRVTLRNGQLWGVEP"

misc_feature 620..3409

/gene="mana"

/note="encodes multidomain beta-1,4-mannanase precursor"

misc_feature		/product="beta-1,4-xylanase XynA"		source	
483..965		/gene="xynA"		1..3509	
misc_feature		/note="Region: thermostabilizing domain"		/organism="Paenibacillus sp. BP-23"	
966..1946		/gene="xynA"		/mol_type="genomic DNA"	
misc_feature		/note="Region: beta-1,4-xylanase catalytic domain"		/strain="BP-23"	
1947..2075		/gene="xynA"		/db_xref="taxon:198119"	
misc_feature		/note="Region: proline-threonine linker"		/country="Spain:Catalonia, Ebro river delta"	
2076..2519		/gene="xynA"		20..3472	
misc_feature		/note="Region: cellulose-binding domain typeIIb"		/gene="cel48C"	
2520..2696		/gene="xynA"		20..25	
misc_feature		/note="Region: proline-threonine linker"		/gene="cel48C"	
2697..3146		/gene="xynA"		39..45	
misc_feature		/note="Region: cellulose-binding domain typeIIb"		/EC_number="3.2.1.91"	
3147..3509		/gene="xynA"		/function="cellulose hydrolysis"	
3510..3509		/gene="xynA"		/codon_start=1	
3510..3509		/gene="xynA"		/transl_table=11	
3510..3509		/gene="xynA"		/product="cellulose 1,4-beta-cellobiosidase"	
3510..3509		/gene="xynA"		/protein_id="CAD32945.1"	
3510..3509		/gene="xynA"		/db_xref="GI:21449824"	
3510..3509		/gene="xynA"		/db_xref="GOA:Q8KKF7"	
3510..3509		/gene="xynA"		/translation="MLKTAAKSLTAMLAGTVMLTGTYGLTGSQTAAHSDQAEIQOA	
3510..3509		/gene="xynA"		DGINEARFLQLYAQLKDPANGVFSAEIGIPYHSETLLSEAPDYGHMSTSEAYSLWL	
3510..3509		/gene="xynA"		ETMYGHTGDSKLEAANDMEKYLIPVNEGDNERQPTMSKYNPNSPATYAAEKPP	
3510..3509		/gene="xynA"		DQPSNLNGQYAAGKDPDLDAELKATYGNQTYLMWLIVDWNWYGNLLNPSTSTY	
3510..3509		/gene="xynA"		VNTFQGEQSVMEAIPIPSQDDK3FGKAGEGFMSLFTKSEQAPSQMRITNAIDADA	
3510..3509		/gene="xynA"		RAVQVLYWAKEMGNNSEYLDKAKMGDFLYKQMYDKYFKQVGSXNGTTPGTGKDS	
3510..3509		/gene="xynA"		NNYLMWYTWGGGLGGDGMARIGASHTQAYQNPVAAVALSDPAGGLIPDPTAK	
3510..3509		/gene="xynA"		ADWNATLKRQLEFYTLQSHHEGAGGATNSIGSVAAPAGVSTFYDMAVQAPVVR	
3510..3509		/gene="xynA"		DPSNTWFGFQAWPLERVAEMTYLLAESGDLTSENFQMAKKVITKWDVTKDYVFVGE	
3510..3509		/gene="xynA"		RPTDAQGYLLNAAGIILGTNAQVATTPAGGFPIPGQEWQGGQDPKNGPSTYTN	
3510..3509		/gene="xynA"		NPNFHAITKDPVQDTCVLGSIYIKALITFFAAGTEAENGTLISAKQEAQKDLAQSLDITAW	
3510..3509		/gene="xynA"		YTDYGVGIWTEERKDYFRFAKEIYLPNWSGTFCQGNITPGTAGVSPDPAKGGNGV	
3510..3509		/gene="xynA"		YTGSDLPALIKQDPAWAYLDNKYTSYNPTTKOWENGAPTFTYHREWSQVDMATAYG	
3510..3509		/gene="xynA"		EFDRLGSSSTFGVQVPSFAGVATAGGDKQVLLSNAAAGATSYSKRATVNGGPVT	
3510..3509		/gene="xynA"		SVATGVSTFTDGLTNGTAYTYVITAVNSAGESAPSAQVATATPOAGTVNGVETLS	
3510..3509		/gene="xynA"		GNAGDAQVNLNWTASTGAATYVQSTGTGAVANLATGLTALTITDTTAVNGTAYNTR	
3510..3509		/gene="xynA"		VYATNAGSGTNNVVLTPMAAPISTGTLEVOYRNGGSSAGNAITPOFNLKNTGTTA	
3510..3509		/gene="xynA"		IDLKVKVRYFTFKDSADMSFCDVAQLGSANVQGSFVAVNPAGKTADTYLLELSFA	
3510..3509		/gene="xynA"		GAGSLAAGAEATGIIQTRFSPKNNWNTFDQSDNYSFNATQTATAGAWTKVTAYQDGAQVMG	
3510..3509		/gene="xynA"		MBP"	
3510..3509		/gene="xynA"		62..166	
3510..3509		/gene="xynA"		/gene="cel48C"	
3510..3509		/gene="xynA"		167..3334	
3510..3509		/gene="xynA"		/gene="cel48C"	
3510..3509		/gene="xynA"		/product="cellulose 1,4-beta-cellobiosidase"	
3510..3509		/gene="xynA"		/EC_number="3.2.1.91"	
3510..3509		/gene="xynA"		3456..3472	
3510..3509		/gene="xynA"		/gene="cel48C"	
3510..3509		/gene="xynA"		sig_peptide	
3510..3509		/gene="xynA"		mat_peptide	
3510..3509		/gene="xynA"		terminator	
3510..3509		/gene="xynA"		ORIGIN	
3510..3509		/gene="xynA"		Alignment Scores:	
3510..3509		/gene="xynA"		Pred. No.:	
3510..3509		/gene="xynA"		Score:	
3510..3509		/gene="xynA"		Percent Similarity:	
3510..3509		/gene="xynA"		Best Local Similarity:	
3510..3509		/gene="xynA"		Query Match:	
3510..3509		/gene="xynA"		DB:	
3510..3509		/gene="xynA"		US-09-917-376-5 (1-88) x PAE488933 (1-3509)	
3510..3509		/gene="xynA"		QY 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnleI 21	
3510..3509		/gene="xynA"		Db 2879 ACGGGAACGCTGGAAGTCAGTATCGCAATGGAGGTCAGTTCAGGTATATCAATT 2938	
3510..3509		/gene="xynA"		QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41	
3510..3509		/gene="xynA"		Db 2939 ACGCGGCAATTAACTCTGAAATAATACAGGTACACGCGCAATTGATCTGAGTAAGGTGAAG 2998	
3510..3509		/gene="xynA"		FEATURES	
3510..3509		/gene="xynA"		Paenibacillus sp. BP-23 cel48C gene for cellulose	
3510..3509		/gene="xynA"		1,4-beta-cellobiosidase.	
3510..3509		/gene="xynA"		Accession AJ488933.1 GI:21449823	
3510..3509		/gene="xynA"		Keywords cel48C gene; cellulose 1,4-beta-cellobiosidase.	
3510..3509		/gene="xynA"		Source Paenibacillus sp. BP-23	
3510..3509		/gene="xynA"		Organism Paenibacillus sp. BP-23	
3510..3509		/gene="xynA"		Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.	
3510..3509		/gene="xynA"		1 Sanchez,M.M., Pastor,F.I.J. and Diaz,P.	
3510..3509		/gene="xynA"		Paenibacillus sp. BP-23 family 48 cellulase. Cloning and	
3510..3509		/gene="xynA"		performance on cellulosic substrates	
3510..3509		/gene="xynA"		Unpublished	
3510..3509		/gene="xynA"		2 (bases 1 to 3509)	
3510..3509		/gene="xynA"		Diaz,P.	
3510..3509		/gene="xynA"		Direct Submission	
3510..3509		/gene="xynA"		Submitted (31-MAY-2002) Diaz P., Microbiology, University of	
3510..3509		/gene="xynA"		Barcelona, Av. Diagonal, 645, 08028-Barcelona, SPAIN	
3510..3509		/gene="xynA"		Location/Qualifiers	
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509		/gene="xynA"			
3510..3509					

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
 Db 2999 GTTCGTATTATTTCAAGAGGAC--AGCGAGCGGACATGAGCTTCTGGTGTGACTAT 3055
 Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
 Db 3056 GCACAGCTTGGCAGTGCACAAATGTGAGCGGACGCTTGTGCTGTAAATCCGCGAAGGCG 3115
 Qy 82 ThrAlaAspThrTyrLeuGln 88
 Db 3116 ACGGAGATACGTATCTCGAG 3136
 RESULT 10
 BSPI33614
 LOCUS BSPI33614 4161 bp DNA linear BCT 09-MAY-2001
 DEFINITION Bacillus sp. BP23 celB gene.
 ACCESSION AJ133614
 VERSION AJ133614.1 GI:4490765
 KEYWORDS celB gene; cellulase; glycosyl hydrolase.
 SOURCE Bacillus sp. BP-23
 ORGANISM Bacillus sp. BP-23
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1
 Pastor,F.I., Pujol,X., Blanco,A., Vidal,T., Torres,A.L. and Diaz,P.
 Molecular cloning and characterization of a multidomain
 endoglucanase from Paenibacillus sp BP-23: evaluation of its
 performance in pulp refining
 Appl. Microbiol. Biotechnol. 55 (1), 61-68 (2001)
 JOURNAL MEDLINE 21129642
 PUBMED 11234960
 REFERENCE 2 (bases 1 to 4161)
 Diaz,P.
 Direct Submission
 Submitted (10-MAR-1999) Diaz P., Microbiology, University of
 Barcelona, Av.Diagonal 645, Barcelona-08028, SPAIN
 JOURNAL Location/Qualifiers
 FEATURES
 source
 /organism="Bacillus sp. BP-23"
 /mol_type="genomic DNA"
 /strain="BP-23"
 /db_xref="taxon:89769"
 398..3525
 /gene="celB"
 398..404
 /gene="celB"
 437..442
 /gene="celB"
 514..519
 /gene="celB"
 532..3525
 /gene="celB"
 /EC_number="3.2.1.4"
 /standard_name="endoglucanase"
 /codon_start=1
 /transl_table=11
 /product="cellulase"
 /protein_id="CAB38941.1"
 /db_xref="GI:4490765"
 /db_xref="GOA:Q92411"
 /db_xref="UniProt/TREMBL:Q92411"
 /translation="MKGSMRVRVAIVTISAGLLAGSFMSMTGLRQADAAAGDHNYAEA
 LOKALYFEARSGPLPANNRVRGDSQMGDAGVDLTCGWYDAGDHVKGFPNHA
 ASATLAWSVVEYADGYEQAGOLEIKDNIWATDYFMKAHTKPNELMGQVAGNTH
 AWGPAEYQMNRRFSFKIDASCPSGDLAAETAAALAAASSIVFADSDPYVSKLQHAH
 ELYNFADYRGKYTDCTIDAAAFYNSMTGYEDLAWGAWLYLATNDNAYLSKALSA
 DRWSTSGSANWPYTWGQWDSKHVGAOILLARTISLNMPEATKFIQSTNSLDYWT
 VCTNGRKYTPGGLAWLDQWGLRYAANAAPISFYVSDVSDVPKSRYPATSOI
 NYLIGDNRQSYVVGYGQSPQPHHRTAHSSWNWNNEDIPANHRHILLYGAWGQPN
 SDQYDDIGDYSVEVATDYNAGFTGALAKNLLYQNHQPLNFPAPPEVKGDFYVE
 AAVRSSGNTYEIRALLNRRSGWPARMGDQLSFYFLDLSVYAGRTVSDVQVTVSS
 SEGATVSPQVVDAAKRIYAITANPNTKIYPGGEGNYRKEVQFRIITGPGQAWNPAND
 PSYONLTGNEVKSYIPYDAGVKVSGEPQTPVAVPAAPAGVQAVAGNSOVALNW
 SASAGASVYTKRAEVSQGPYTTVAAGVNGLYTNTGLTNGKTYTYVVTAVNSAGESP

ASVQSGMPOAATTVPAGVTLTSGTAGNNQNMMLSWTAASGASSYTVQRAVAGTGTDDVA
 TGLAVLNTDTTALNGTSYSYRAAVNASGOTLSNVVMLTTPSGPVPSTGTLVQVRSRG
 GSGNSNAVTPOFLKNTGTQALDLSVTKIRYFTKDGTELSFPCWCDVAQVGSANVOG
 MFVAVNPAGKTADTVVEISPTSGAGSLAAGAETGVIQTRFSKNNSAFDQSNDSYDA
 SKTAPAAWKVTAYQGNTOVWGLEP"
 532..633
 /gene="celB"
 ORIGIN
 Alignment Scores:
 Pred. No.: 7,03e-15 Length: 4161
 Score: 204.50 Matches: 38
 Percent Similarity: 67.82% Conservative: 21
 Best Local Similarity: 43.68% Mismatches: 27
 Query Match: 43.60% Indels: 1
 DB: 1 Gaps: 1
 US-09-917-376-5 (1-88) x BSPI33614 (1-4161)
 Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
 Db 3067 ACAGGTACGCTTGGTGCAGTATCGAAGCGGGGTTGAGGCAATCCAGCAATGCAGTT 3126
 Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThr 41
 Db 3127 ACTCGCAATTCATCTCAAGAAATACCGGCACACAGCGGATGATCTGAGTACGGTGA 3186
 Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
 Db 3187 ATTCGTTATTACTTTTACCAAGAC---GGCAGGAGGAGCTCTCTCTGTTGTTGATTAT 3243
 Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
 Db 3244 GCGAGGTGGGCTCAGCGAAGTACAGCAATGTTTGTAGCAGTGAATCCGCGAAGGT 3303
 Qy 82 ThrAlaAspThrTyrLeuGln 88
 Db 3304 ACAGCGGATACGTATGTGAG 3324
 RESULT 11
 BSPI3142
 LOCUS E35142 2029 bp DNA linear PAT 18-JUN-2001
 DEFINITION Truncated cellulase composition.
 ACCESSION E35142
 VERSION E35142.1 GI:13018967
 KEYWORDS JP 1999221086-A/44.
 SOURCE synthetic construct
 ORGANISM other sequences; artificial sequences.
 REFERENCE 1 (bases 1 to 2029)
 AUTHORS Paiji,A., Petaer,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,
 Hyu,M. and Daian,P.W.
 TITLE Truncated cellulase composition
 JOURNAL Patent: JP 1999221086-A 44 17-AUG-1999;
 CLARIANT INTERNATIONAL LTD
 COMMENT OS Artificial Sequence
 PN JP 1999221086-A/44
 PD 17-AUG-1999
 PF 21-SEP-1998 JP 1998283606
 PI 19-SEP-1997 US 08/932571
 PI PAIJI ANDERSON,PETAER L BAGUKUISUTO,ROY M DANIEL, PI
 GURAHAMU K PARINTON,
 PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONIORISU WILLIAM
 PC C12N15/09,C11D3/386,C12N1/21,C12N9/42//C12N1/21,C12R1:19) , PC
 (C12N9/42,C12R1:19),C12N15/00
 CC
 FH Key Location/Qualifiers
 FT source 1..2029 /organism='Artificial Sequence'.
 FT Location/Qualifiers
 1..2029
 /organism="synthetic construct"
 /mol_type="genomic DNA"

/db_xref="taxon:32630"

ORIGIN

Alignment Scores: 3.78e-14 Length: 2029
Pred. No.: 195.50 Matches: 38
Score: 195.50
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.68% Indels: 1
DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x E35142 (1-2029)

```
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1 ATGGGAAGTGGTGTGAAGTACTGTACAAAGCAATGAGCAAGTGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 61 ATAGCGCGTGGTTTAAAGATAGTGAATGGAGGCAGCAGCAGTGTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 121 AAGATAAGACTACTGGTACACAGTGGTGTGACAAAGCCACAGAGTGCAGTA---TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 238 AGTGAGCGGATATTACCTGGAG 261
```

RESULT 12

E35143 2029 bp DNA linear PAT 18-JUN-2001
LOCUS
DEFINITION Truncated cellulase composition.
ACCESSION E35143
VERSION E35143.1 GI:13018968
KEYWORDS JP 1999221086-A/45.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 2029)
AUTHORS Paiji,A., Peters,L.B., Roy,M.D., Gurahamu,K.F., Moreland,D.G.,
Hyu,M. and Dalian,P.W.
TITLE Truncated cellulase composition
JOURNAL Patent: JP 1999221086-A 45 17-AUG-1999;
COMMENT CLARIANT INTERNATIONAL LTD
OS Unidentified
PN JP 1999221086-A/45
PD 17-AUG-1999
PF 21-SEP-1998 JP 1998283606
PI 19-SEP-1997 US 08/932571
PR PAIJI ANDERSON,PETAER L BAGUKUISUTO,ROY M DANIEL, PI
GURAHAMU K FARINTON,
PI MORELAND DAVID GIBUSU,HYU MORGAN,DAIAN PURATONIORISU WILLIAM
PC C12N15/09,C11D3/386,C12N1/21,C12N9/42//C12N1/21,C12R1:19), PC
(C12N9/42,C12R1:19),C12N15/00

CC
FH Key Location/Qualifiers
FT source 1..2029
FT Location/Qualifiers
1..2029 /organism='Unidentified'.
/db_xref="taxon:32630"

FEATURES

source
1..2029
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN

Alignment Scores: 3.78e-14 Length: 2029
Pred. No.: 195.50 Matches: 38
Score:

Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.68% Indels: 1
DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x E35143 (1-2029)

```
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1 ATGGGAAGTGGTGTGAAGTACTGTACAAAGCAATGAGCAAGTGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 61 ATAGCGCGTGGTTTAAAGATAGTGAATGGAGGCAGCAGCAGTGTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 121 AAGATAAGACTACTGGTACACAGTGGTGTGACAAAGCCACAGAGTGCAGTA---TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 238 AGTGAGCGGATATTACCTGGAG 261
```

RESULT 13
AF078038S5 6005 bp DNA linear BCT 11-FEB-2001
LOCUS
DEFINITION Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 6 gene, partial
cds; and CelE gene, complete cds.
ACCESSION AF078042
VERSION AF078042.1 GI:12743878
KEYWORDS
SEGMENT
SOURCE
ORGANISM
5 of 5
Caldicellulosiruptor sp. Tok7B.1
Caldicellulosiruptor sp. Tok7B.1
Bacteria; Firmicutes; Clostridia; Clostridiales;
Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE 1 (bases 1 to 6005)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Bergquist,P.L.
TITLE Multidomain and multifunctional glycosyl hydrolases from the
extreme thermophile Caldicellulosiruptor isolate Tok7B.1
JOURNAL Curr. Microbiol. 40 (5), 333-340 (2000)
MEDLINE 20171169
PUBMED 10706665
REFERENCE 2 (bases 1 to 6005)
AUTHORS Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
Williams,D.P. and Bergquist,P.L.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Biological Sciences, Macquarie University,
Sydney, NSW 2109, Australia
FEATURES
source
1..6005 Location/Qualifiers
/organism="Caldicellulosiruptor sp. Tok7B.1"
/mol_type="genomic DNA"
/isolates="Tok7B.1"
/db_xref="taxon:80339"
1..506
/notes="ORF6; multidomain"
/codon_start=3
/transl_table=11
/product="glycosyl hydrolase 6"
/protein_id="AAK06393.1"
/db_xref="GI:12743884"
/translation="GQPDWTNGTYTGNPNLHVKIDYGTDLGITASLANALLYSAAT
KEYGVSDEAKNLAKELDRMNLVIRDDKGLSAPKRGDYKRFEEQVYIPAGWTGKM
PNGDVIKSGVKFIDIRSKYQKDPDQWLVSAYNAGEAPEPRFHRFWAQCDIAIANTY
EILFGNQ"
misc_feature
1..503
/note="Region: family 48 glycosyl hydrolase domain"

CDS	617..5872	Db	2708	TGTTTAAAGATAGTAGTAATGGAGCGCAGCAGCTGTTGATCTTTAGCAGGTTAAGATAAGA	2767
	/note="multidomain glycosyl hydrolase"	Qy	44	TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63
	/codon_start=1	Db	2768	TACTGTTACACAGCTGATGGTGACCAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG	2824
	/evidence=experimental	Qy	64	MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla	83
	/transl_table=11	Db	2825	ATAGGCGAAGCAATGTGACATTCATTTTGTGAAGCTTTAGCAGCGGAGTGAGTGGAGCG	2884
	/product="CeleB"	Qy	84	AspThrTyrLeuGln	88
	/db_xref="GI:12743885"	Db	2885	GATTATTACCTGGAG	2899
	/protein_id="AAK06394.1"	RESULT 14			
	/translation="MQSMKAIKRVSITALLVLTSLCFPGIMPVKAAGCTVNYGREA	E35100			
	LQKTIMEFVFMQSKLFSWNRNRGSGDLDKDVGLDLTGGHWDAGDHVKFNLPMS	LOCUS	E35100	6416 bp	DNA
	YSASMLGVAATVEYKDLFIHLNLEIANDYFVKQLEPSKIVYQVGDPTVDH	DEFINITION	E35100	Truncated cellulase composition.	
	NFWGPAEYMQMRPAYKCDLSNPASSVVAETAFLASVAVIKERNQKAAASYLQHA	ACCESSION	E35100		
	DLFEADTTRADGATGATGTYTSGGFTDILGAAVMLYIATNDSYLTAKAEELMSEY	VERSION	E35100.1	GI:13018925	
	ANGNTWTQDWDVRYGTLIMLKITGKELVGAVERNLHDWDRITVTPKGMAYLTG	KEYWORDS	JP 199221086-A/2.		
	WGSRYATTAFLACVADGSCDKNKYLNFPAKSOIDYALGSGTSFVVGRTNY	SOURCE	unidentified		
	POPHRHNASHSWANSMKIPEYHRHILYALVGFGSGDDSYNDITDYQNEVACDYN	ORGANISM	unclassified.		
	AGIVGALAKMTLYGGPIDDFAKETPTNDDEIFVESFGNSQGPNEYTEVLSIYNRT	REFERENCE	1 (bases 1 to 6416)		
	GWPRVTDKLSFKYFIDLELIIQAGYSPDVVKVDTYIEGKISGPVYMDKRNIIYY	AUTHORS	Paiji, A., Petaer, L.B., Roy, M.D., Gurahamu, K.F., Moreland, D.G.,		
	LVDPSGTIKYPEGVEHKKQAFKISVPQGYPMWPTNDPSYKGLTSOLEKNKYAAVD	TITLE	Truncated cellulase composition		
	NNNLVWGLPECAATSTAPTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT	JOURNAL	Patent: JP 199221086-A 2 17-AUG-1999;		
	YKNETSASTGSIKRPWKIVNGSGSSVDLSRVKIRYMYVDGDKPQSAWCDMAOIGAS	COMMENT	CLARIANT INTERNATIONAL LTD		
	NVTNFVKLSGSGADYILEVFGSGAGQCPKDGDIQVRFNKNWNSYNQADDW		OS Unknown		
	SWLQSMYNGNAKTYLVDGLVWGPEPGATPAPTATPTPTPTPTPTPTPTPTPTPT		PN JP 199221086-A/2		
	TPTVATPTAPTASPVGSYTPESYSGALKVWYANGNLSPTNVLNPKIKIENVTG		PD 17-AUG-1999		
	TAVDLRVKRYWTIDGEATQSVSVASSINPAYIDVKFVKLGANAGGADYVVEIGFK		PF 21-SEP-1998 JP 1998283606		
	SGVLAAGQSTKEIRLSIOKSGSYNQSNDSYRSANSYIENKVTGYIDDVLVWGR		PI 19-SEP-1997 US 08/932571		
	EPGRNAQIKWYANGNLGSMTNVLPKIKIENVTGTAVDLSRVKRYMYTIDSGEATQS		PI PAIJI ANDERSON, PETAER L BAGUKUISUTO, ROY M DANIEL, PI		
	VSVTSSINPAYIDVKFVKLGANAGADYIEFGSGAGVLAAGQKEIRLSIQGS		GURAHAMU K FARINTON,		
	GSYNQSNDSYRSATGYIENKVTGYIDGAIWGREPSRGTKPAGGVTPPTAPTST		PI MORELAND DAVID GIBUSU, HYU MORGAN, DAIAN PURATONITOTISU WILLIAM		
	PT		PC C12N15/09, C11D3/386, C12N1/21, C12N9/42/(C12N1/21, C12R1:19), PC		
	VLGTGFKGYVMAARWGDGSGYAOAAVNIYLNKDGKSGRYGTCVSAETTVENMPVY		(C12N9/42, C12R1:19), C12N15/00		
	ASIEGDDSTVHILINRNYDRKLKAEIKMNRVYTGGEIYGFDPDSTSSQIRKMGVLS				
	NIQNNTTIEVNPNTVHVLTSSK"				
	725..2053				
	/note="Region: family 9 glycosyl hydrolase domain"	FEATURES			
	2054..2545	source			
	/note="type IIic cellulose binding site"	CC			
	2627..3106	Key			
	/note="type IIb cellulose binding site"	Location/Qualifiers			
	3224..4195	FT			
	/note="duplicated type IIb cellulose binding site"	FT			
	4298..5869	source			
	/note="Region: family 44 glycosyl hydrolase domain"	Location/Qualifiers			
	2537..2626	1..6416			
	/note="PT-linker"	/organism="unidentified"			
	3107..3223	/mol_type="genomic DNA"			
	/note="PT-linker"	/db_xref="taxon:32644"			
	4196..4295				
	/note="PT-linker"				
	ORIGIN				
	Alignment Scores:				
	Pred. No.:	2.13e-13	Length:	6416	
	Score:	194.50	Matches:	38	
	Percent Similarity:	61.18%	Conservative:	14	
	Best Local Similarity:	44.71%	Mismatches:	32	
	Query Match:	41.47%	Indels:	1	
	DB:	1	Gaps:	1	
	US-09-917-376-5 (1-88) x AF078038S5 (1-6005)				
Qy	4 GlyVallysValGlnTyrIlyAsnAsnAspSerAlaProGlyAspAsnGlnIleLysPro	23			
Db	2648 GGTGTGAAGGTAAGTGTACAAACAATGAGCAAGTGGCAGCACAGGTTCTATAAGCGCG	2707			
Qy	24 GlyLeuGlnLeuValAsnThrClySerSerValAspLeuSerThrValThrValArg	43			

	2725 TGGTTTAAAGATAGTAGTAATGGAGCGCAGCAGCTGTTGATCTTTAGCAGGTTAAGATAAGA	2784			
Qy	44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla	63			
Db	2785 TACTGTTACACAGCTGATGGTGACCAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG	2841			

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 06:17:10 ; Search time 135.381 Seconds
(without alignments)
3847.935 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKQYKXNDSPGDNQ.....IRASFGSVNPATPTDYLQ 88

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame-p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool/US09917376/runat 04102005.164327 26889/app query.fasta 1.2524
-DB=N_Geneseq_16Dec04 -QWTF=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CGN 1 1 1867 @runat 04102005.164327 26889 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	469	100.0	2289	8 ABZ77633	Abz77633 Nucleotid
2	469	100.0	2289	10 ADJ38292	Adj38292 A. cellu
3	466	99.4	3687	8 ABZ77634	Abz77634 Nucleotid
4	466	99.4	3687	12 ADH36637	Adh36637 DNA encod
5	462	98.5	2869	10 ADD22922	Add22922 Acidother

6	462	98.5	2869	10 ABZ77632	Abz77632 Nucleotid
7	462	98.5	2869	12 ADOS2314	Ados2314 Acidother
8	436	93.0	3365	8 ABZ76162	Abz76162 A. cellu
9	409	87.2	3364	10 ADF75896	Adf75896 Acidother
10	231.5	49.4	2600	2 AAQ15178	Aaq15178 Portion o
11	195.5	41.7	2029	2 AAX55660	Aax55660 DNA seque
12	195.5	41.7	2029	6 AAD26568	Aad26568 Active ce
13	194.5	41.5	6415	2 AAX55662	Aax55662 DNA seque
14	194.5	41.5	6416	6 AAD26526	Aad26526 Active ce
15	192.5	41.0	11706	2 AAX55661	Aax55661 DNA seque
16	192.5	41.0	11707	6 AAD26525	Aad26525 Active ce
17	190	40.5	1527	13 ADT48400	Adt48400 Acterial
18	190	40.5	2177	11 ADOS5906	Ados5906 Bacillus
19	188	40.1	1434	6 AAL41028	Aal41028 CMCase ge
20	188	40.1	1488	6 AAL41025	Aal41025 CMCase ge
21	188	40.1	2510	6 ABK53202	Abk53202 Bacillus
22	186	39.7	2175	2 AAQ49820	Aaq49820 NK-1 cell
23	182.5	38.9	2977	2 AAQ13001	Aaq13001 Endol gen
24	165	35.2	501	10 ADG14263	Adg14263 cellulose
25	165	35.2	717	10 ADG14257	Adg14257 Cellulose
26	165	35.2	1438	3 AAZ45336	Aaz45336 DNA encod
27	165	35.2	1482	2 AAX90978	Aax90978 DNA encod
28	165	35.2	1482	2 AAZ31562	Aaz31562 Pectate l
29	165	35.2	5562	2 AAT86625	Aat86625 C. thermo
30	163	34.8	1500	12 ADJ35111	Adj35111 DNA encod
31	157	33.5	1314	6 ABK73393	Abk73393 Bacillus
32	157	33.5	2602	12 ADG32259	Adg32259 DNA encod
33	155.5	33.2	4059	5 AAH75059	Aah75059 Nucleotid
34	150.5	32.1	4056	5 AAH75060	Aah75060 Nucleotid
35	130.5	27.8	4452	8 ACA27429	Aca27429 Prokaryot
36	126.5	27.0	1303	10 ADC27473	Adc27473 Fusion pr
37	126.5	27.0	1747	10 ADC27474	Adc27474 Fusion pr
38	126.5	27.0	3489	10 ADC27475	Adc27475 Fusion pr
39	120.5	25.7	1775	2 AAQ13003	Aaq13003 Endo3 gen
40	114	24.3	486	2 AAQ72917	Aaq72917 Cellulose
41	114	24.3	486	2 AAV74072	Aav74072 C. cellul
42	114	24.3	499	2 AAX24930	Aax24930 Clostridi
43	114	24.3	507	5 AAD11042	Aad11042 Clostridi
44	114	24.3	768	4 AAF86248	Aaf86248 DNA seque
45	114	24.3	768	6 ABK52403	Abk52403 E. coli c

ALIGNMENTS

RESULT 1
ABZ77633
ID ABZ77633 standard; DNA; 2289 BP.
XX
AC ABZ77633;
XX
DT 03-JUN-2003 (first entry)
XX
DE Nucleotide sequence of the ManA polypeptide.
XX
KW ManA; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
KW food; feed; paper pulp; biofuel; manase; gene; ss.
XX
OS Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
FT CDS 1..2289
FT /*tag= a
FT /product= "ManA"
XX
PN WO2003012110-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023819.
XX
PR 28-JUL-2001; 2001WO-US023819.
XX
XX (WIDE) MIDWEST RES INST.
PA


```

XX  Ding S, Adney WS, Vinzant TB, Himmel ME;
PI  WPI; 2003-248182/24.
XX  P-PSDB; ABP73022.
XX  Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT  cellulolyticus, useful for reducing hemicellulose in a starting material,
PT  for processing of food, and as bulking agents in food stuffs.
XX  Example 1; Page 23; 46pp; English.
XX  The present sequence encodes ManaA, a thermostable mannanase A polypeptide
CC  derived from Acidothermus cellulolyticus. ManaA is a member of the
CC  glycoside hydrolase family of enzymes. ManaA is useful for reducing
CC  hemicellulose in a starting material to simpler carbohydrate units, and
CC  ultimately to sugars which are useful in the food, feed, paper pulp, and
CC  biofuels industries. It is useful for the processing of food and in food
CC  stuffs as bulking agents, and for the degradation of mannanase. ManaA is
CC  also useful to raise polyclonal and monoclonal antibodies that are useful
CC  in purifying ManaA, or detecting ManaA polypeptide expression, and as well
CC  as reagent tools for characterizing the molecular actions of ManaA
CC  polypeptides
XX  SQ  Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      5,36e-48      Length:      2289
Score:          469.00      Matches:      88
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              8          Gaps:      0

US-09-917-376-5 (1-88) x ABZ77633 (1-2289)

QY  1  ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB  1363  GTGTGGGTGGGTGAAGTGCAGTACAGAACATGATTGGCGCCGGGTGATACACG 1422

QY  21  IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
DB  1423  ATCAAAACGGGTCTCCAGTTGGTGAATACGGGGTCTGTCGGTGGATTGTTCGACGGTG 1482

QY  41  ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB  1483  ACGTGTGGTACTGGTTTCCCGGGATGGTGGGTCTGTCGACACTGGTGTACAACTGTGAC 1542

QY  61  TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB  1543  TGGCGCGCGATGGGGTGTGGGAAATATCCCGCTCTGTTCCGCTCGTGAACCCGCGGACG 1602

QY  81  ProThrAlaAspThrTyrLeuGln 88
DB  1603  CCGACGGCGGACACCTACCTGCAG 1626

RESULT 2
ADU38292
ID  ADU38292 standard; DNA; 2289 BP.
XX
XX  AC  ADU38292;
XX
XX  DT  06-MAY-2004 (first entry)
XX
XX  DE  A. cellulolyticus mannanase ManaA DNA.
XX  KW  mannanase A; catalytic domain GH5; carbohydrate binding domain III;
XX  KW  carbohydrate binding domain II; food processing; foodstuff;
XX  KW  bulking agent; ManaA; ds; Gene.
XX  OS  Acidothermus cellulolyticus.
XX
XX  FH  Key      Location/Qualifiers

```

```

CDS
FT  1..2289
FT  /*tag= a
FT  /EC number= "3.2.1.78"
FT  /product= "Mannanase A"
XX
XX  US2003119093-A1.
XX  26-JUN-2003.
XX  28-JUL-2001; 2001US-00917378.
XX  28-JUL-2001; 2001US-00917378.
XX  (DING/) DING S.
XX  PA  (ADNE/) ADNEY W S.
XX  PA  (VINZ/) VINZANT T B.
XX  PA  (HIMM/) HIMMEL M E.
XX
XX  Ding S, Adney WS, Vinzant TB, Himmel ME;
XX  WPI; 2003-863435/80.
XX  P-PSDB; ADJ38291.
XX
XX  New composition comprising a purified mannanase A peptide, comprising a
PT  catalytic domain GH5, carbohydrate binding domain III or carbohydrate
PT  binding domain II, useful in food processing and in food stuffs as
PT  bulking agents.
XX
XX  Claim 11; SEQ ID NO 2; 18pp; English.
XX
XX  The invention relates to a composition comprising a purified mannanase A
CC  peptide, having a catalytic domain GH5, carbohydrate binding domain III
CC  or carbohydrate binding domain II. The composition is useful in food
CC  processing and in foodstuffs as bulking agents. The present sequence
CC  represents the A. cellulolyticus mannanase ManaA DNA.
XX
XX  SQ  Sequence 2289 BP; 463 A; 700 C; 688 G; 438 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      5,36e-48      Length:      2289
Score:          469.00      Matches:      88
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              10         Gaps:      0

US-09-917-376-5 (1-88) x ADJ38292 (1-2289)

QY  1  ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB  1363  GTGTGGGTGGGTGAAGTGCAGTACAGAACATGATTGGCGCCGGGTGATACACG 1422

QY  21  IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
DB  1423  ATCAAAACGGGTCTCCAGTTGGTGAATACGGGGTCTGTCGGTGGATTGTTCGACGGTG 1482

QY  41  ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB  1483  ACGTGTGGTACTGGTTTCCCGGGATGGTGGGTCTGTCGACACTGGTGTACAACTGTGAC 1542

QY  61  TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB  1543  TGGCGCGCGATGGGGTGTGGGAAATATCCCGCTCTGTTCCGCTCGTGAACCCGCGGACG 1602

QY  81  ProThrAlaAspThrTyrLeuGln 88
DB  1603  CCGACGGCGGACACCTACCTGCAG 1626

RESULT 3
ABZ77634
ID  ABZ77634 standard; DNA; 3687 BP.
XX
XX  AC  ABZ77634;

```


XX 03-JUN-2003 (first entry)
XX Nucleotide sequence of the GuxA polypeptide.
XX GuxA; cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;
XX detergent; pulp processing; paper processing; feed processing; textile;
XX gene; ss.
XX Acidothermus cellulolyticus.
XX Key Location/Qualifiers
XX CDS 1..3687
XX FT /*tag= a
XX FT /product= "GuxA"
XX WO2003012109-A1.
XX 13-FEB-2003.
XX 28-JUL-2001; 2001WO-US023817.
XX 28-JUL-2001; 2001WO-US023817.
XX (WIDE) MIDWEST RES INST.
XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX WPI; 2003-239526/23.
XX P-PSDB; ABP73029.
XX Novel thermal tolerant GuxA polypeptide derived from Acidothermus
XX cellulolyticus, useful for reducing cellulose in a starting material, and
XX for the conversion of biomass to biofuels and biofuel additives.
XX Example 1; Page 23-24; 47pp; English.
XX The present sequence encodes a GuxA polypeptide. GuxA is thermostable
XX cellulase, and is a member of the glycoside hydrolase family of enzymes.
XX GuxA is useful for reducing cellulose in a starting material such as
XX agricultural biomass to sugars. This is useful in biofuel production.
XX GuxA is also useful in the conversion of biomass to biofuels and biofuel
XX additives, in detergents, pulp and paper processing, food and feed
XX processing, and in textile processes. GuxA is also useful for raising
XX polyclonal and monoclonal antibodies that are useful in purifying GuxA,
XX or detecting GuxA polypeptide expression, as well as reagent tools for
XX characterizing the molecular actions of GuxA polypeptides
XX
XX Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 2,33e-47 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.36% Indels: 0
DB: Gaps: 0
US-09-917-376-5 (1-88) x ABZ77634 (1-3687)
QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAenGln 20
Db 1750 GTGTGGGTGGGTGAAGGTGCAGTACAGAACATGATTCGGCGCGGTGATACCCAG 1809
QY 21 IleValProGlyLeuGlnLeuValAlaThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGGTCTCCAGTTGGTGAATACCGGTCTGTCGGTGGATTGTGCAGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGTGGCGGTACTGGTTTACCCCGGATGGTGGGTCTGCACACTGGTGTACACTGTGAC 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAenProAlaThr 80

Db 1930 TGGCGCGGATGGGTGTGGGAATATCCGCCCTCGTTTCGGTCAACCCGCGCAG 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACTGCGAG 2013
RESULT 4
ADH36637
ID ADH36637 standard; DNA; 3687 BP.
XX
XX AC ADH36637;
XX DT 11-MAR-2004 (first entry)
XX DE DNA encoding Acidothermus cellulolyticus glycoside hydrolase, GuxA.
XX KW Thermal tolerant cellulase; glycoside hydrolase; GuxA;
XX KW carbohydrate degradation; cellulose; agricultural biomass;
XX KW municipal solid waste; thermostable; gene; ds.
XX OS Acidothermus cellulolyticus.
XX Key Location/Qualifiers
XX CDS 1..3687
XX FT /*tag= a
XX FT /product= "GuxA"
XX US2003104522-A1.
XX 05-JUN-2003.
XX 28-JUL-2001; 2001US-00917383.
XX 28-JUL-2001; 2001US-00917383.
XX (DING/) DING S.
XX (ADNE/) ADNEY W S.
XX (VINZ/) VINZANT T B.
XX (HIMM/) HIMMEL M E.
XX (DECK/) DECKER S R.
XX Ding S, Adney WS, Vinzant TB, Himmel ME, Decker SR;
XX WPI; 2004-106451/11.
XX P-PSDB; ADH36636.
XX Novel isolated thermostable GuxA polypeptide useful for detecting
XX polynucleotide encoding GuxA, assessing carbohydrate degradation activity
XX of GuxA, reducing cellulose in starting material e.g., agricultural
XX biomass.
XX Claim 21; SEQ ID NO 2; 20pp; English.
XX The present invention relates to the isolation of a thermal tolerant
XX cellulase from Acidothermus cellulolyticus. The cellulase is a member of
XX the glycoside hydrolase family and is designated GuxA. Also disclosed are
XX methods of producing and using GuxA. The GuxA polypeptide is useful for
XX the detection of a polynucleotide encoding GuxA. The polypeptide sequence
XX is also useful for assessing the carbohydrate, e.g. cellulose,
XX degradation activity of GuxA. GuxA is useful for reducing cellulose in a
XX starting material such as agricultural biomass or municipal solid waste.
XX The polypeptide molecule of GuxA is thermostable. The present sequence
XX encodes A. cellulolyticus glycoside hydrolase, GuxA.
XX Sequence 3687 BP; 725 A; 1173 C; 1171 G; 618 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 2,33e-47 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 98.86% Mismatches: 0
Query Match: 99.36% Indels: 0

DB: 12 Gaps: 0

US-09-917-376-5 (1-88) x ADR36637 (1-3687)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20

Db 1750 GTGTCGGGTGGCTGAAGTGCAGTACAAACAATGATTCGGCGCCGGGTGATAACACAG 1809

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40

Db 1810 ATCAACCGGGTCTCCAGTGTGTGATACCGGGTCTGTCGTGGATTTGTGACGGTG 1869

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

Db 1870 ACGGTGGGTACTGTTCAACCGGGATGTTGGTGTGTCGACACTGTTGTACAACTGTGAC 1929

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80

Db 1930 TGGCGCGGATGGGGTGTGGGAATATCCGCGCTCGTTCGGTCCGTAACCCGCGACG 1989

QY 81 ProThrAlaAspThrTyrLeuGln 88

Db 1990 CCGACGGCGACACCTACCTCGAG 2013

RESULT 5

ADD22922

ID ADD22922 standard; DNA; 2869 BP.

XX AC ADD22922;

XX 15-JAN-2004 (first entry)

DE Acidothermus cellulolyticus avicelase AvIII DNA.

XX AvIII; cellulose reduction; agricultural biomass; municipal solid waste;

KW glycoside hydrolase; avicelase; ds; gene.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

FT CDS 1..2869

FT /*tag= a

FT /product= "AvIII"

XX US2003108988-A1.

XX 12-JUN-2003.

XX 18-OCT-2002; 2002US-00155400.

XX 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

PA (ADNE/) ADNEY W S.

PA (VINZ/) VINZANT T B.

PA (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-810853/76.

DR P-PSDB; ADD22921.

XX New isolated thermal tolerant avicelase polynucleotide useful for

PT detection of a polynucleotide encoding AvIII and for reducing cellulose

PT in a starting material, e.g. municipal solid waste.

XX Claim 17; SEQ ID NO 2; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a

CC thermostable AvIII polypeptide. The polynucleotide is useful for

CC detection of a polynucleotide encoding AvIII. The polynucleotide is

CC useful for reducing cellulose in a starting material which involves

CC administering to the starting material, e.g. agricultural biomass or

CC municipal solid waste, a polypeptide molecule of the polynucleotide. The

CC method further comprises administering a second polypeptide molecule

CC chosen from the glycoside hydrolase family of proteins. The present

CC sequence represents DNA encoding the Acidothermus cellulolyticus

CC avicelase AvIII.

XX SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores: 5.3e-47 Length: 2869

Pred. No.: 462.00 Matches: 86

Score: 100.00% Conservatives: 2

Percent Similarity: 97.73% Mismatches: 0

Best Local Similarity: 98.51% Indels: 0

Query Match: 10 Gaps: 0

DB: 2869

US-09-917-376-5 (1-88) x ADD22922 (1-2869)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20

Db 2605 GTGTCGGGTGGGTGAAGTGCAGTATAAGAAATAATGATTCGGCGCCGGGTGATAATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40

Db 2665 ATCAACCGGGTTCGAGGTGTGATACCGGGTCTGTCGTGGATTTGTGACGGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

Db 2725 ACGGTGGGTACTGTTCAACCGGGATGTTGGTGTGTCGACACTGTTGTACAACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80

Db 2785 TGGCGCGGATGGGGTGTGGGAATATCCGCGCTCGTTCGGTCCGTAACCCGCGACG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88

Db 2845 CCGACGGCGACACCTACCTCGAG 2868

RESULT 6

ABZ77632

ID ABZ77632 standard; DNA; 2869 BP.

XX AC ABZ77632;

XX 03-JUN-2003 (first entry)

DE Nucleotide sequence of the avicelase AvIII.

XX Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;

KW detergent; pulp processing; paper processing; feed processing; textile;

KW cellulose; gene; ss.

XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers

FT CDS 1..2869

FT /*tag= a

FT /partial

FT /product= "AvIII"

FT /transl_except= (pos:2869,aa:Xaa)

FT /note= "Xaa is an unspecified residue"

XX WO2003012090-A2.

PN 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023818.

XX 28-JUL-2001; 2001WO-US023818.

XX (MIDE) MIDWEST RES INST.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

Db 2785 TGGCGCGGATCGGGTGTGGGAATATCGCGCTCGTTCGGTGAACCCGCGGACG 2844
 |||...|||

Qy 81 ProThrAlaAspThrTyrLeuGln 88
 2845 CCGACGGCGGACACCTACCTGCAG 2868
 |||...|||

RESULT 8
 ABZ76162
 ID ABZ76162 standard; DNA; 3365 BP.
 XX
 AC ABZ76162;
 XX

DT 29-MAY-2003 (first entry)
 XX
 DE A. cellulolyticus Gux1 protein encoding DNA.
 XX

KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;
 KW biofuel; detergent; pulp; paper processing; feed processing; textile;
 KW cellulase; gene; ds.
 XX

OS Acidothermus cellulolyticus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..3365
 FT /tag= a
 FT /product= "Gux1"
 FT /trans_except= (pos: 682..683, aa: Pro)
 FT /note= "this codon has an apparent one nucleotide
 basepair deletion which alters the reading frame"

XX
 PN WO2003012095-A1.
 XX
 PD 13-FEB-2003.
 XX
 PF 28-JUL-2001; 2001WO-US023820.
 XX
 PR 28-JUL-2001; 2001WO-US023820.
 XX
 PA (MIDE) MIDWEST RES INST.
 XX

PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;
 PI Lantz McCarter S;
 XX
 PI WPI; 2003-300494/29.
 DR P-PSDB; ABP71656.
 XX

PT New thermal tolerant Gux1 peptide having specified amino acid sequence,
 PT useful in the degradation of cellulose to biofuels.
 XX
 PS Disclosure; Page 22-23; 44pp; English.
 XX

CC The invention relates to a thermal tolerant Gux1 peptide from A.
 CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside
 CC hydrolase family and comprises a catalytic domain GH48, carbohydrate
 CC binding domain type III, and a carbohydrate binding domain type II. The
 CC polypeptide is useful in the degradation of cellulose into biofuel, or
 CC for conversion of biomass to biofuel additives. It is used in detergents,
 CC pulp and paper processing, food and feed processing, and in textile
 CC processing. It can also be used alone or in combination with other
 CC cellulase or glycoside hydrolases. The novel polypeptide generates
 CC alternative cellulase enzymes capable of assising in the commercial-
 CC scale processing of cellulose to sugar for use in biofuel production. The
 CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding
 CC DNA
 XX
 SQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.11e-43 Length: 3365
 Score: 436.00 Matches: 81
 Percent Similarity: 97.62% Conservative: 1
 Best Local Similarity: 96.43% Mismatches: 2

Query Match: 92.96% Indels: 0
 DB: 8 Gaps: 0

US-09-917-376-5 (1-88) x ABZ76162 (1-3365)

Qy 5 VallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
 ::|||...|||

Db 112 CTCAAAGCGCAGTATAAGAACCAATGATTCGGCGCGAGTGACCAACCAACCGGGT 171
 |||...|||

Qy 25 LeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyr 44
 |||...|||

Db 172 CTCCAGTTGGTGAATATACCGGGTCGTCGTCGGTGGATTTGTCGACGGTGACGGTCCGGTAC 231
 |||...|||

Qy 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaMet 64
 |||...|||

Db 232 TGGTTCACCCGGGATGGTGGTCTGTCGACACTGGTGTACAACTGTGACTGGGCGCGATG 291
 |||...|||

Qy 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84
 |||...|||

Db 292 GGGTGTGGGAATATCCGGCCTCGTTGGCTCGGTGAACCCGCGGACCGCGCGGAC 351
 |||...|||

Qy 85 ThrTyrLeuGln 88
 |||...|||

Db 352 ACCTACCTGCAG 363
 RESULT 9
 ADF75896
 ID ADF75896 standard; DNA; 3364 BP.
 XX
 AC ADF75896;
 XX
 DT 26-FEB-2004 (first entry)
 XX
 DE Acidothermus cellulolyticus Gux1 DNA.
 XX

KW Gux1; gene; ds; thermal tolerant; cellulase; glycoside hydrolase;
 KW exoglucanase; cellulose; sugar; biofuel production.
 XX
 OS Acidothermus cellulolyticus.
 XX

FH Key Location/Qualifiers
 FT CDS 1..3364
 FT /tag= a
 FT /product= "Gux1 protein"
 FT /note= "This polynucleotide sequence contains translation
 exceptions"

XX
 PN US2003096342-A1.
 XX
 PD 22-MAY-2003.
 XX

PF 28-JUL-2001; 2001US-00917384.
 XX
 PR 28-JUL-2001; 2001US-00917384.
 XX

PA (ADNE/) ADNEY W S.
 PA (DING/) DING S.
 PA (VINZ/) VINZANT T B.
 PA (HIMM/) HIMMEL M E.
 PA (DECK/) DECKER S R.
 PA (MCCA/) MCCARTER S L.
 XX

PI Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR, McCarter SL;
 PI WPI; 2003-863404/80.
 DR P-PSDB; ADF75895.
 XX

XX Novel thermal tolerant cellulase of glycoside hydrolase family,
 PT comprising catalytic domain, first and second carbohydrate binding
 PT domain, isolated from Acidothermus cellulolyticus, useful in cellulose
 PT degradation.
 XX
 PS Claim 21; SEQ ID NO 2; 20pp; English.

XX This invention relates to a novel thermal tolerant cellulase, a member of
CC the glycoside hydrolase family. Specifically, it refers to an
CC exoglucanase identified as Gux1, and isolated from *Acidothermus*
CC *cellulolyticus*, which comprises the catalytic domain GH48, and the
CC carbohydrate binding domains type II and type III. The present invention
CC describes Gux1 compositions that are useful for reducing cellulose in a
CC starting material (or biomass). As such, these enzymes are capable of
CC assisting in commercial scale processing of cellulose to sugar, for use
CC in biofuel production. Furthermore, heterologous cellulase proteins (or
CC fusion proteins) can be generated with specific desired properties
CC including thermal stability and partial or complete resistance to extreme
CC pH inactivation, proteolytic, solvent, detergent or chaotropic agent
CC inactivation. This polynucleotide sequence is the *Acidothermus*
CC *cellulolyticus* Gux1 DNA of the invention.

XX SQ Sequence 3364 BP; 669 A; 1110 C; 1048 G; 536 T; 1 U; 0 Other;

Alignment Scores:
Pred. No.: 2.52e-40 Length: 3364
Score: 409.00 Matches: 80
Percent Similarity: 96.43% Conservative: 1
Best Local Similarity: 95.24% Mismatches: 3
Query Match: 87.21% Indels: 1
DB: 10 Gaps: 0

US-09-917-376-5 (1-88) x ADF75896 (1-3364)

QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
DB 112 CTCGAAGCGCAGTATAGACATGATTCGGCGCGAGTGACCAACAGATCAACCGGGT 171
QY 25 LeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyr 44
DB 172 CTCAGTTGGTGAATACCGGGTCGTGCGTGGGAUTG-TCGACGCTGACGGTGGCGTAC 230
QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64
DB 231 TGGTTTCCACCGGGATGGTGGGTGCGACACTGGGTGTAACCTGTGACTGGCGCGGATG 290
QY 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84
DB 291 GGGTGTGGGAATATCCGGCCCTCGTTCCGCTCGGTGAACCCGGCGAGCGCGGAC 350
QY 85 ThrTyrLeuGln 88
DB 351 ACCTACCTGCAG 362

RESULT 10
AAQ15178
ID AAQ15178 standard; DNA; 2600 BP.

XX AAQ15178;

XX 27-AUG-2003 (revised)
XX 25-MAR-2003 (revised)
XX 17-MAR-1992 (first entry)

XX Portion of pAEC-1 contg. cellulase AE-1 gene.

XX Detergents; pharmaceuticals; deinking; carboxymethylcellulose; ss.

XX *Aeromonas* sp.

XX Key Location/Qualifiers

XX CDS 192..2540

XX /*tag= a

XX /product= "cellulase AE-1"

XX JP03251174-A.

XX 08-NOV-1991.

XX

PF 28-FEB-1990; 90JP-00045465.
XX
PR 28-FEB-1990; 90JP-00045465.
XX
PA (OJIP) OJI PAPER CO.
XX
XX WPI; 1991-373412/51.
DR P-PSDB; AAR15625.
XX
XX Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs - of opt.
PT pH when carboxymethylcellulose is used as substrate.
XX
XX Claim 3; Fig 2; 8pp; Japanese.
XX
XX The plasmid, pAEC 1, was prepd. ligating chromosomal DNA contg. the gene
CC (obtd. from *Aeromonas*) into pUC18. The gene can be used to produce
CC recombinant enzymes which is used for the effective utilis- ation of
CC biomass resources and the mfr. of pharmaceuticals and foodstuffs, and
CC also for the detergent and deinking of waste paper. (Updated on 25-MAR-
CC 2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ Sequence 2600 BP; 687 A; 649 C; 700 G; 564 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.11e-18 Length: 2600
Score: 231.50 Matches: 42
Percent Similarity: 66.67% Conservative: 16
Best Local Similarity: 48.28% Mismatches: 28
Query Match: 49.36% Indels: 1
DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAQ15178 (1-2600)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 2082 AGCGGGGATTTGGCGGTTCAGTACAAACCGGAGATACGAATCGACGCCGACACAGTTC 2141
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThr 41
DB 2142 AAGCCTCATTTCAATATCTGTGAACAGGGGGTTCGGGAGTGCCTTTAAGCGAGCTGAGT 2201
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrIeuValTyrAsnCysAspTrp 61
DB 2202 CTGCGATATATTTCACAGCTGAC---GGCAATGACCACTGCAATACAAATTGTGACTGG 2258
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 2259 GCCATGTAGTGTCTCCAACTGCAACCTGAAACGGGGCTTTCGTGAAAATGAATCCGGGCAACGCG 2318
QY 82 ThrAlaAspThrTyrIeuGln 88
DB 2319 AACGCGGATACCTACTTGGAG 2339

RESULT 11
AAQ55660
ID AAX55660 standard; DNA; 2029 BP.

XX AAX55660;

XX 30-JUL-1999 (first entry)

XX DNA sequence encoding truncated cellulase Cel E3/B5.

XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
XX Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
XX cotton-containing fabric; stonewashing; ss.

XX Unidentified.

XX EP921188-A2.

XX 09-JUN-1999.

XX


```

PF 15-SEP-1998; 98EP-00810919.
XX
XX 19-SEP-1997; 97US-00932571.
XX
XX (CLRN ) CLARIANT FINANCE BVI LTD.
XX
XX Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;
PI Morgan H, Williams DP;
XX
XX WPI; 1999-315403/27.
DR P-PSDB; AAY13494.
DR
XX New truncated cellulase proteins, useful in detergents and for producing
PT 'stonewashed' denim.
XX
XX Disclosure; Page 41-42; 65pp; English.
XX
XX The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from
CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3
CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751
CC and the stability region extends from amino acid E482 to G635 in the
CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or remove
CC staining, backstaining or graying, for use on cellulosic materials
CC including cotton-containing fabrics. They are especially useful for
CC preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using non-
CC truncated cellulase compositions
XX
XX Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.59e-14 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.68% Indels: 1
DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAX55660 (1-2029)
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1 ATGGGAAGTGGTGAAGGACTCTGTACAAGAACATGACACAAAGTGGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
DB 61 ATAAGCCCGTGGTTTAAGATAGTGAATGGAGCGCAGCAGTGTGTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
DB 121 AAGATAAGATACTGGTACACAGTGGATGGTGCACAAAGCCACAGAGTGGCGGTA---TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
DB 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
DB 238 ACTGGAGCGGATATTACCTGGAG 261

RESULT 12
AAD26568
ID AAD26568 standard; DNA; 2029 BP.
XX
XX AC AAD26568;

```

```

XX 26-MAR-2002 (first entry)
DT Active cellulase hybrid protein, E3/B5 DNA.
DE
XX Active cellulase protein; alkaliphilic; textile processing; proteinase;
KW detergent additive; stonewashed appearance; cotton-containing denim;
KW CelB5; thermophilic; commercial detergent; E3/B5 hybrid protein; ds.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
FH 1. .1851
FT CDS /*tag= a
FT FT /product= "E3/B5 hybrid protein"
XX
XX US6294366-B1.
XX
XX 25-SEP-2001.
XX
XX 19-AUG-1998; 98US-00136574.
XX
XX 19-SEP-1997; 97US-00932571.
XX (CLRN ) CLARIANT FINANCE BVI LTD.
XX
XX Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
PI Morgan H, Williams DP;
XX
XX WPI; 2002-081780/11.
DR P-PSDB; AAE16325.
XX
XX New cellulase active protein, useful in textile processing or commercial
PT detergents, e.g. for improving the feel or appearance of cotton-
PT containing fabrics, is stable under conditions of alkaline pH and
PT elevated temperatures.
XX
XX Disclosure; Col 71-74; 61pp; English.
XX
XX The present invention relates to a cellulase active protein, which is
CC substantially free of proteinases of native thermophilic and
CC alkaliphilic origin, where the cellulase active protein consists of the
CC CelB5 amino acid sequence. The cellulase active protein is useful for
CC treating cellulosic materials including cotton-containing fabrics, as
CC detergent additives. The cellulase active protein is also useful for
CC improving the feel and/or appearance of cotton-containing fabrics, for
CC removing surface fibers from cotton-containing knits or for imparting
CC stonewashed appearance to cotton-containing denims. The present proteins
CC are stable under condition of alkaline pH and elevated temperatures, thus
CC suitable for textile processing and in commercial detergents. The present
CC sequence is E3/B5 hybrid protein DNA
XX
XX Sequence 2029 BP; 650 A; 329 C; 520 G; 530 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.59e-14 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.68% Indels: 1
DB: 6 Gaps: 1

US-09-917-376-5 (1-88) x AAD26568 (1-2029)
QY 1 ValSerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
DB 1 ATGGGAAGTGGTGAAGGACTCTGTACAAGAACATGACACAAAGTGGCAGCAGGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
DB 61 ATAAGCCCGTGGTTTAAGATAGTGAATGGAGCGCAGCAGTGTGTGATCTTAGCAGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60

```



```

Db 121 AAGATAAGTACTGGTACACAGTGTGTGACAGCCACAGAGTGGGTA---TGTGAC 177
      ::::::::::::::::::::|
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
      ::::::::::::::::::::|
Db 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGAGTG 237
      ::::::::::::::::::::|
QY 81 ProThrAlaAspThrTyrLeuGln 88
      ::::::::::::::::::::|
Db 238 AGTGGAGCGGATTATTACCTGGAG 261
      ::::::::::::::::::::|

RESULT 13
AAX55662
ID AAX55662 standard; DNA; 6415 BP.
XX
AC AAX55662;
XX
DT 30-JUL-1999 (first entry)
XX
DE DNA sequence encoding truncated cellulases.
XX
KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
KW cotton-containing fabric; stonewashing; ss.
XX
OS Unidentified.
XX
PN BP921189-A2.
XX
PD 09-JUN-1999.
XX
PF 15-SEP-1998; 98EP-00810919.
XX
PR 19-SEP-1997; 97US-00932571.
XX
PA (CLRN ) CLARIANT FINANCE BVI LTD.
XX
PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;
PI Morgan H, Williams DP;
XX
DR WPI; 1999-315403/27.
DR P-PSDB; AAY13493.
XX
XX New truncated cellulase proteins, useful in detergents and for producing
PT 'stonewashed' denim.
XX
PS Disclosure; Page 24-25; 65pp; English.
XX
CC The invention relates to a recombinant cellulase active protein free of
CC proteinases of native thermophilic and alkaliphilic origin, comprising
CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
CC length sequences, or functional equivalents. Cel B5 extends from amino
CC acid A1011 to F1424 or K1425 or N1426, and Cel B4/5 extends from amino
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from
CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3
CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751
CC and the stability region extends from amino acid E482 to G635 in the
CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
CC enzymes are useful in laundry detergent compositions to prevent or remove
CC staining, backstaining or graying, for use on cellulosic materials
CC including cotton-containing fabrics. They are especially useful for
CC preventing redeposition of colorant during stonewashing, and for
CC processing of textiles where cellulose breakdown is required. The new
CC truncated enzymes show reduced redeposition of dye compared to using non-
XX truncated cellulase compositions
XX
SQ Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 0 U; 36 Other;

Alignment Scores:
Pred. No.: 2.66e-13 Length: 6415
Score: 194.50 Matches: 38
Percent Similarity: 61.18% Conservative: 14

```

```

Best Local Similarity: 44.71% Mismatches: 32
Query Match: 41.47% Indels: 1
DB: 2 Gaps: 1

US-09-917-376-5 (1-88) x AAX55662 (1-6415)

QY 4 GlyVallyVsValGlnTyrIlyAsnIleArgAlaSerPheGlyAspAsnGlnIleLysPro 23
      ::::::::::::::::::::|
Db 2664 GGTGTGAAGGTACTGTACAGAACCAATGACAGTGCAGCGACAGGTTCTTATAGGCCG 2723
      ::::::::::::::::::::|
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
      ::::::::::::::::::::|
Db 2724 TGGTTTAAGATAGTGAATGGAGCGACAGCAGTGTTGATCTTAGCAGGGTTAAGATAGA 2783
      ::::::::::::::::::::|
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
      ::::::::::::::::::::|
Db 2784 TACTGTACACAGTGTGACAGCCACAGAGTGGGTA---TGTGACTGGGCACAG 2840
      ::::::::::::::::::::|
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
      ::::::::::::::::::::|
Db 2841 ATAGGGGCAAGCAATGTGACATTCAATTTTGTGAAGCTTAGCAGCGAGTGTGAGCG 2900
      ::::::::::::::::::::|
QY 84 AspThrTyrLeuGln 88
      ::::::::::::::::::::|
Db 2901 GATTATTACCTGGAG 2915
      ::::::::::::::::::::|

RESULT 14
AAX55626
ID AAX55626 standard; DNA; 6416 BP.
XX
AC AAX55626;
XX
DT 26-MAR-2002 (first entry)
XX
DE Active cellulase protein, celE gene.
XX
KW Active cellulase protein; alkaliphilic; textile processing; proteinase;
KW detergent additive; stonewashed appearance; cotton-containing denim;
KW CelB5; thermophilic; commercial detergent; celE gene; ds.
XX
OS Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 634..5889
XX FT /*tag= a
XX FT /product= "CelE protein"
XX FT misc_feature 748..2538
XX FT /*tag= c
XX FT /product= "DNA encoding E1/2 protein"
XX FT misc_feature 748..2076
XX FT /*tag= b
XX FT /product= "DNA encoding E1 protein"
XX
PN US6294366-B1.
XX
XX 25-SEP-2001.
XX
XX 19-AUG-1998; 98US-00136574.
XX
XX 19-SEP-1997; 97US-00932571.
XX
XX (CLRN ) CLARIANT FINANCE BVI LTD.
XX
XX Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
XX Morgan H, Williams DP;
XX
XX WPI; 2002-081780/11.
XX P-PSDB; AAE16324.
XX
XX New cellulase active protein, useful in textile processing or commercial
XX PT detergents, e.g. for improving the feel or appearance of cotton-
XX PT containing fabrics, is stable under conditions of alkaline pH and
XX PT elevated temperatures.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 07:43:21 ; Search time 846.19 Seconds
(without alignments)
3958.515 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKQYKKNDSAPGDNQ.....IRASFGSVNPATPTADTVLQ 88

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09917376/runat_04102005.164328.26911/app.query.fasta_1.2524
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOFL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @CNC 1 1 12009 @runat_04102005.164328.26911 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	81	17.3	584	7	CF480372
2	79.5	17.0	754	5	BQ047524
3	78.5	16.7	768	6	CB293085
4	77	16.4	906	5	BQ722991
5	76	16.2	376	1	AA043716
6	76	16.2	606	2	BE362891
7	76	16.2	641	6	CD226529
8	76	16.2	657	2	BE362820
9	76	16.2	676	7	CF487649

10	76	16.2	691	4	BG739392
11	76	16.2	724	8	BZ616323
12	76	16.2	814	4	BI218457
13	76	16.2	905	8	BZ545807
14	76	16.2	934	9	CG270741
15	76	16.2	1001	9	CG270748
16	75.5	16.1	657	2	BF513128
17	75.5	16.1	1041	9	CL974250
18	75.5	16.1	1058	7	CK231562
19	75.5	16.1	2091	3	HSMB00308
20	75	16.0	706	7	CK352071
21	74.5	15.9	465	2	AW491464
22	74.5	15.9	613	6	CA595709
23	74.5	15.9	650	8	BZ894843
24	74	15.8	302	7	CF076729
25	74	15.8	463	9	BX959940
26	74	15.8	74	4	BI664037
27	73.5	15.7	261	2	BF875908
28	73.5	15.7	752	7	CN187748
29	73.5	15.7	798	7	CN187749
30	73.5	15.7	816	6	CB293980
31	73.5	15.7	819	6	CB293084
32	73.5	15.7	894	7	CK938508
33	73.5	15.7	1039	3	AY539888
34	73.5	15.7	1039	4	BG334904
35	73.5	15.7	1237	5	BQ065850
36	73	15.6	357	7	CF078232
37	73	15.6	360	6	C64738
38	73	15.6	593	7	CF622108
39	73	15.6	776	9	BX209426
40	73	15.6	861	8	BH319803
41	73	15.6	894	9	CG077853
42	73	15.6	924	9	CG077856
43	72.5	15.5	380	4	BG816375
44	72.5	15.5	450	1	AL799805
45	72.5	15.5	533	2	AW963120

ALIGNMENTS

RESULT 1
CF480372
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CF480372 584 bp mRNA linear EST 08-SEP-2003
POL1_65_H07_b1_A002 Pollen Sorghum bicolor cDNA clone
POL1_65_H07_A002 3', mRNA sequence.
CF480372
CF480372.1 GI:34509241
EST.
Sorghum bicolor (sorghum)
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 584)
Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,
Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J.,
Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and
Pratt,L.H.
EST database from Sorghum: pollen
Unpublished (2003)
Other ESTs: POL1_65_H07_g1_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpatratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude


```

449 GAA-----CCTCCTTCGCGAGATAGTTACCTT 475
Db
RESULT 3
CB293085      768 bp  mRNA      linear      EST 28-FEB-2003
LOCUS        UCRCS01_05cb12_g1 Washington Navel orange cold acclimated flavedo &
DEFINITION   albedo cDNA library Citrus sinensis cDNA clone UCRCS01_05cb12, mRNA
sequence.
CB293085
ACCESSION    CB293085.1 GI:28618542
VERSION      CB293085.1
KEYWORDS     Citrus sinensis
SOURCE       Citrus sinensis
ORGANISM     Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
1 (bases 1 to 768)
Close,T.J., Collura,K., Fenton,R.D., Feuerbacher,O., Kim,H.R.,
Kudrna,D., Wanmaker,S., Wing,R. and Yu,Y.
Development of EST Resources and New Genetic Markers for California
Citrus
Unpublished (2003)
Contact: Timothy Close
Department of Botany & Plant Sciences, University of California
Riverside, CA, 92521-0124
Tel: 9097873318
Fax: 9097874437
Email: timothy.close@ucr.edu
Seq primer: T3.

FEATURES             Location/Qualifiers
     source
     1..768
         /organism="Citrus sinensis"
         /mol_type="mRNA"
         /cultivar="Washington navel"
         /db_xref="taxon:2711"
         /clone="UCRCS01_05cb12"
         /tissue_type="kind containing flavedo and albedo"
         /dev_stage="Mature fruit"
         /lab_host="E. coli TJC121"
         /clone_lib="Washington Navel orange cold acclimated
         flavedo & albedo cDNA library"
         /note="Vector: Lambda Uni-ZAP XR, excised phagenid;
         Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
         field at University of California, Riverside Agricultural
         Operations since 1983. The scion was Washington Navel
         orange and the rootstock Carizzo Citrange. Tissue from
         mature fruit was harvested at mid-day in January 2002
         during a cold spell, when pre-dawn temperatures were
         approximately -2 to 2 degree C. Approximately 2 cm median
         sections of the rind were excised in the field from
         several fruits, then wrapped in aluminum foil and frozen
         quickly in dry ice. Total RNA was extracted using a phenol
         extraction procedure described in J. Japanese Soc. Hort.
         Sci. (1996) 64:809-814. Poly(A) RNA was purified, a cDNA
         library was made, and 1 million primary lambda cDNA clones
         were in vivo excised to give a population of pBluescript
         SK(-) phagenids. All steps to this point were performed in
         the TJ Close lab at the University of California,
         Riverside (Fenton). Phagenids were plated, plasmid DNA
         purified, cDNA clones archived, and DNA sequences
         determined bi-directionally using an ABI3530 at the
         Arizona Genomics Institute, University of Arizona
         (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu).
         Chromatogram files were transmitted to UC Riverside (by
         Yu), then processed at UC Riverside (by Wanmaker) using
         the HarVest pipeline (http://harvest.ucr.edu) to remove
         vector and cloning oligo sequences and various
         contaminants, and to trim to a high quality region.
         Sequences that retained a phred 17 region of at least 100
         bases were deposited to GenBank."

ORIGIN
Alignment Scores: 58.7 Length: 906
Pred. No.: 30.7 Length: 768
Score: 78.50 Matches: 20
Percent Similarity: 57.69% Conservative: 10
Best Local Similarity: 38.46% Mismatches: 15
Query Match: 16.74% Indels: 7
DB: 6 Gaps: 2

US-09-917-376-5 (1-88) x CB293085 (1-768)
QY 3 GlyGlyValLysValGlnTyrLysAsnAsnSerAlaProGlyAspAsnGlnIleLys 22
Db 612 GCGCGCATCAAC-----AAGACACACACAGCAGCAATAATAAAGTAAT 662
QY 23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42
Db 663 AATAATACGACGTTGTACTAGTGTAGCGTAGCGTTGTGTCAAGTGGCGTGACG--- 719
QY 43 ArgTyrTrpPheThrArgAspGlyGlySerSerThr 54
Db 720 -----TTTTCGAGAGCGGAGGAGCAGT 746

RESULT 4
BQ722991      906 bp  mRNA      linear      EST 16-JUL-2002
LOCUS        BQ722991
DEFINITION   AGENCOURT 8103733 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6191052 5', mRNA sequence.
ACCESSION    BQ722991.1 GI:21861888
VERSION      BQ722991.1
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 906)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13590 row: n column: 13
High quality sequence stop: 512.

FEATURES             Location/Qualifiers
     source
     1..906
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="IMAGE:6191052"
         /sex="male"
         /tissue_type="sympathetic trunk"
         /dev_stage="adult, 16 yr"
         /lab_host="DH10B"
         /clone_lib="Lupski sympathetic trunk"
         /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:
         NotI; Site 2: SalI; cDNA made by oligo-dT priming.
         Directionally cloned using the following adaptors:
         5'-TCGACCCACGCGTCG-3' and
         5'-GACTAGTTCTAGTCGCGAGCGGCCCGCCCT(15)-3'. Size selected >
         1 kb for average insert length 1.9 kb. This is a primary
         library, non-amplified. Library constructed by Life
         technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
         College of Medicine); available through Life
         Technologies."

ORIGIN
Alignment Scores: 58.7 Length: 906
Pred. No.: 30.7 Length: 906
Score: 78.50 Matches: 20
Percent Similarity: 57.69% Conservative: 10
Best Local Similarity: 38.46% Mismatches: 15
Query Match: 16.74% Indels: 7
DB: 6 Gaps: 2

```


Score: 77.00 Matches: 16
 Percent Similarity: 49.15% Conservatives: 13
 Best Local Similarity: 27.12% Mismatches: 26
 Query Match: 16.42% Indels: 4
 DB: 5 Gaps: 2

US-09-917-376-5 (1-88) x BQ722991 (1-906)

Qy 7 ValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLysProGlyLeuGln 26
 Db 548 CTGGCTATACCAATAAAATCTCCAGGAGGCCACCGTGAGTA-----ATGAAG 598

Qy 27 LeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrpPhe 46
 Db 599 GCTGTAAAGTGGGAATGAAGAATATGAGTTGGAAGCCTCTTCCAGCACTACTGGTAC 658

Qy 47 ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
 Db 659 TCCCGG---GGCGGCATCGGCCAGCTCTTACACCTGCATCTGGGCGAGTGTGGA 712

RESULT 5

AA043716 376 bp mRNA linear EST 11-MAY-1997

LOCUS zk61f10.s1 Soares_pregnant_uterus NBHPU Homo sapiens cDNA clone
 DEFINITION IMAGE:487339 3' similar to gb:L10240 BASIGIN PRECURSOR (HUMAN);
 mRNA sequence.

AA043716

AA043716.1 GI:1521725

EST.

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (Bases 1 to 376)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.P., Chiapelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
 Trevas, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
 and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

9704478

8889549

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1400 Std Error: 0.00

Seq primer: -40M13 fwd. from Amersham

High quality sequence stop: 212.

FEATURES

source

1..376

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="GDB:3761053"

/db_xref="taxon:9606"

/clone="IMAGE:487339"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="Soares_pregnant_uterus_NBHPU"

/note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I;

Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo (dT) primer [5'

AACGTGAGAATTCGGCGCGCTTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by M. Fatima Bonaldo."

ORIGIN

Alignment Scores: 25.2 Length: 376
 Pred. No.: 76.00 Matches: 20
 Score: 43.42% Conservatives: 13
 Percent Similarity: 26.32% Mismatches: 23
 Best Local Similarity: 16.20% Indels: 20
 Query Match: 1 Gaps: 3
 DB: 1

US-09-917-376-5 (1-88) x AA043716 (1-376)

Qy 17 GlyAspAsnGlnLysProGlyLeuGlnLeuValAsnThrGlySerSerValAsp 36
 Db 64 GGGGGTTCACGAGGAGCCCTGGT---GCTATAGANAGCCGCCCTCTCCAGGGGTGAG 120

Qy 37 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerThr----- 54
 Db 121 CTGGGGAGGCTT---CTGCGGTCTGGAGTCCCGCGATGGCGCCAGTTCCCGAGCAAC 177

Qy 55 -----LeuValTyrAsn 58

Db 178 CCCCTCCAGAGCTGCCCGGATGCACAGACAGGAGGGGCTTGGGAGTGACTTGAGGC 237

Qy 59 CysAspTrpAlaAlaMetGlyCysGlyAsnLysArgAlaSerPheGly 74
 Db 238 TGTACGGGGTGCCTCGGGTGTGGCAAGTGAAGTCTCTGTGGGC 285

RESULT 6

BE362891

LOCUS

DEFINITION

BE362891

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Jul 20, 2000 this sequence version replaced gi:9304448.

Contact: Cordomier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20. Three-prime sequences, which are obtained with PolyTMix or

T7 sequencing primer, are presented as the reverse complement.

Seq primer: PolyTMix

High quality sequence start: 15

High quality sequence stop: 594

POLYA=No.

Location/Qualifiers

1..606

/organism="Sorghum bicolor"

/mol_type="mRNA"

/db_xref="taxon:4558"

/clone_lib="Dark Grown 1 (DGL)"

/note="Organ: 5-day-old dark-grown seedlings; Vector:

Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was

made from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."

ORIGIN

Alignment Scores: 46.7 Length: 606
Pred. No.: 76.00 Matches: 23
Score: 44.15% Conservative: 15
Percent Similarity: 26.74% Mismatches: 38
Best Local Similarity: 16.20% Indels: 10
Query Match: 2 Gaps: 4
DB:

US-09-917-376-5 (1-88) x BE362891 (1-606)

QY 5 ValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnLleLysProGly 24
DB 208 ATAGCAATCAGCGAATCAATCAATCGGTCCGGAGGAGCGTGTGTCGGCCGGGC 267
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
DB 268 -----CGGCCGGGACAGAACCTCGAACTTCCAAACGATCGTGCAGAGAAA 315
QY 44 ---TyrTrp---PheThrArgAspGlySerSerThrLeuValThrValThrValArg 61
DB 316 CCATGTTGGGGATCTGCAGGAGGAGGTTTTCAGCAACATGGAATATATATGATGAA 375
QY 62 AlaAlaMetGlyCysGlyAsnLle-----ArgAlaSerPheGlySerValAsnPro 78
DB 376 GCTTGTGTGAAGTGTGTGACGTGTTCTTTGGAGACGCTGCTTGTCTGTGTGTCGG 435
QY 79 AlaThrProThrAlaAsp 84
DB 436 AGCTTCAGTCTCGTGAT 453

RESULT 7
CD226529 641 bp mRNA linear EST 21-MAY-2003
LOCUS CCC1_46_E07_b1 A007 Callus culture/cell suspension Sorghum bicolor
DEFINITION cDNA clone CCC1_46_E07_A007 3', mRNA sequence.
CD226529
ACCESSION CD226529.1 GI:30969963
VERSION
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 641)
AUTHORS Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S.,
Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Rathore,K.,
Eastman,A. and Pratt,L.H.
An EST database from Sorghum: callus culture and cell suspension
Unpublished (2003)
Other_ESRs: CCC1_46_E07_g1_A007
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

TITLE Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
JOURNAL the Human Genome Center, University of Tokyo Institute of Medical
COMMENT Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.
Seq primer: Sug3 (CGACCTGCAGCTCGAGCAC)

POLYA=Yes. Location/Qualifiers

FEATURES
source 1. .641

/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="RTx430"
/db_xref="taxon:4558"
/clone="CCC1_46_E07_A007"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Callus culture/cell suspension"
/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The
library was prepared from a mixture of polyA+ RNA from
callus culture tissue and cells in suspension culture.
Double-stranded cDNA was cloned unidirectionally into
different DraIII sites of the pME18S-FL3 vector (5-prime
DraIII site is CACGTGTG, 3-prime DraIII site is
CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Alignment Scores: 50.2 Length: 641
Pred. No.: 76.00 Matches: 23
Score: 44.19% Conservative: 15
Percent Similarity: 26.74% Mismatches: 38
Best Local Similarity: 16.20% Indels: 10
Query Match: 6 Gaps: 4
DB:

US-09-917-376-5 (1-88) x CD226529 (1-641)

QY 5 ValLysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnLleLysProGly 24
DB 149 ATAGCAATCAGCGAATCAATCAATCGGTCCGGAGGAGCGTGTGTCGGCCGGGC 208
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
DB 209 -----CGGCCGGGACAGAACCTCGAACTTCCAAACGATCGTGCAGAGAAA 256
QY 44 ---TyrTrp---PheThrArgAspGlySerSerThrLeuValThrValThrValArg 61
DB 257 CCATGTTGGGGATCTGCAGGAGGAGGTTTTCAGCAACATGGAATATATATGATGAA 316
QY 62 AlaAlaMetGlyCysGlyAsnLle-----ArgAlaSerPheGlySerValAsnPro 78
DB 317 GCTTGTGTGAAGTGTGTGACGTGTTCTTTGGAGACGCTGCTTGTCTGTGTGTCGG 376
QY 79 AlaThrProThrAlaAsp 84
DB 377 AGCTTCAGTCTCGTGAT 394

RESULT 8
BE362820 657 bp mRNA linear EST 20-JUL-2000
LOCUS DG1_B9_C02.g2_A002 Dark grown 1 (DG1) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.
BE362820
ACCESSION BE362820.1 GI:9304377
VERSION
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 657)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and
Pratt,L.H.
An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence

is 20.
Seq primer: PolyTmix
High quality sequence start: 89
High quality sequence stop: 651
POLYA=No.

FEATURES
source
1. .657
/location/Qualifiers
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

FEATURES
source
1. .676
/location/Qualifiers
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTX623"
/db_xref="taxon:4558"
/clone="POL1_45_C08_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Pollen"
/note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from pollen at the late vacuolated-vacuolated stage of development. Pollen was harvested from greenhouse-grown panicles of sorghum line BTX623. Panicles were removed from the flag leaf prior to emergence, when no detectable amylase is present in pollen of male-fertile lines. This stage represents pollen collected from anthers about 8-14 days prior to anthesis. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Alignment Scores:
Pred. No.: 53.7 Length: 676
Score: 76.00 Matches: 23
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match: 16.20% Indels: 10
DB: 7 Gaps: 4

US-09-917-376-5 (1-88) x CF487649 (1-676)

QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
::: ::: :
Db 214 ATAGCAATCAGCGGAACTCAATGAATCCGTCGCGGAGGAGCCTGTGTCGCGCGGC 273
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
::: ::: :
Db 274 -----CGCGCGGAGAGAACCTCCAACTTCCAAACGATCGGTGCAAGGAAA 321
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTirp 61
::: ::: :
Db 322 CCATCGTGGGATCTCGAGGAGAGAGGTTTGAGCAACATGGAATATATATATGATGAA 381
QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
::: ::: :
Db 382 GCTTGTGTGAAGTGTGGTGTGACCTTGTCTTTGGAGACGCTGTCTTGTCTGTGTCGCG 441
QY 79 AlaThrProThrAlaAsp 84
::: ::: :
Db 442 AGCTTGCAGTCTGTGTAT 459

RESULT 10
BG739392
LOCUS
DEFINITION
EM1_82_D05_g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
ACCESSION
BG739392
VERSION
BG739392.1 GI:14089081

FEATURES
source
1. .657
/location/Qualifiers
/organism="Sorghum bicolor"
/mol_type="mRNA"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

ORIGIN
Alignment Scores:
Pred. No.: 51.8 Length: 657
Score: 76.00 Matches: 23
Percent Similarity: 44.19% Conservative: 15
Best Local Similarity: 26.74% Mismatches: 38
Query Match: 16.20% Indels: 10
DB: 2 Gaps: 4

US-09-917-376-5 (1-88) x BE362820 (1-657)

QY 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
::: ::: :
Db 204 ATAGCAATCAGCGGAACTCAATGAATCCGTCGCGGAGGAGCCTGTGTCGCGCGGC 263
QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
::: ::: :
Db 264 -----CGCGCGGAGAGAACCTCCAACTTCCAAACGATCGGTGCAAGGAAA 311
QY 44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTirp 61
::: ::: :
Db 312 CCATCGTGGGATCTCGAGGAGAGAGGTTTGAGCAACATGGAATATATATATGATGAA 371
QY 62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
::: ::: :
Db 372 GCTTGTGTGAAGTGTGGTGTGACCTTGTCTTTGGAGACGCTGTCTTGTCTGTGTCGCG 431
QY 79 AlaThrProThrAlaAsp 84
::: ::: :
Db 432 AGCTTGCAGTCTGTGTAT 449

CF487649 676 bp mRNA linear EST 08-SEP-2003
POL1_45_C08_b1_A002 Pollen Sorghum bicolor cDNA clone
CF487649
CF487649.1 GI:34516518
EST.
Sorghum bicolor (sorghum)
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 676)
Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J., Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J., and Pratt,L.H.
EST database from Sorghum: pollen
Unpublished (2003)
Other_Ests: POL1_45_C08_g1_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860


```
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 691)
AUTHORS Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
TITLE An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 50
High quality sequence stop: 691
POLYA=No.

FEATURES             source
                    1..691
                    /organism="Sorghum bicolor"
                    /mol_type="mRNA"
                    /db_xref="taxon:4558"
                    /clone_lib="Embryo 1 (EM1)"
                    /note="Organ: Embryos germinated for 24 hr; Vector:
                    pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
                    EcoRI; The library was made from poly-A RNA in the cloning
                    vector lambda ZAP II. Clones to be sequenced were
                    prepared by mass excision."

ORIGIN
Alignment Scores:      Length:      691
Pred. No.:            55.3          Matches:      23
Score:                76.00         Conservative: 15
Percent Similarity:   44.1%         Mismatches:   38
Best Local Similarity: 26.74%        Indels:       10
Query Match:          16.20%        Gaps:         4
DB:                   4

US-09-917-376-5 (1-88) x BG739392 (1-691)

QY      5 VallysValGlnTyrIlysnAnsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24
      ::: ::: :::::|||||
DB      265 ATAGCAATCAGCCGAAACTCAATGAATCCGTCGGGAGGAGCCTGTGCTGGCCGGGC 324

QY      25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg--- 43
      :::::|||||:::|||||:::
DB      325 -----CGGCGCGGACAGAACCTCGAACTTCAACGATCGTGCAGGAAA 372

QY      44 ---TyrTrp---PheThrArgAspGlyGlySerSerThrLeuValThrAsnCysAspTrp 61
      :::::|||||:::|||||:::
DB      373 CCATGTGGGGATCTCGCAGGAGGAGGAGGTTTGAGCAACATGGAATATATATGATGAA 432

QY      62 AlaAlaMetGlyCysGlyAsnIle-----ArgAlaSerPheGlySerValAsnPro 78
      :::::|||||:::|||||:::
DB      433 GCTTGTGTGAAGTGTGTGACGTGTGTTCTTTGGAGACGTTGTCTTGTCTGTGTGTCGG 492

QY      79 AlaThrProThrAlaAsp 84
      :::::|||||
DB      493 AGCTTGACGCTGTGTGAT 510

RESULT 11
BZ616323/c
LOCUS
DEFINITION 724 bp DNA linear GSS 16-JAN-2003
1958e08.g1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
1958e08 5', genomic survey sequence.
ACCESSION BZ616323
```

```
VERSION BZ616323.1 GI:27766878
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 724)
AUTHORS Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N.,
Katzburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zutavern, I., McCombie, W.R. and Martienssen, R.A.
TITLE Genomic shotgun sequences from Zea mays (methyl-filtered)
JOURNAL Unpublished (2002)
COMMENT Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: ig58 row: e column: 08
Seq primer: -21M3UnivRev
Class: shotgun
High quality sequence stop: 728.

FEATURES             source
                    1..724
                    /organism="Zea mays"
                    /mol_type="genomic DNA"
                    /cultivar="B73"
                    /db_xref="taxon:4577"
                    /clone="ig58e08"
                    /lab_host="DH5a"
                    /clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
                    /note="Organ: Immature ears; Site 1: Xba I; Site 2: Xba I;
                    The vector was digested with XbaI and one nucleotide was
                    added by fill in the recessive 3' end. The genomic DNA
                    was nebulized, end repaired, adaptor ligated and size
                    fractionated using sephadex. The resulting fragments were
                    between 0.8 and 3 kb and were cloned into the vector (.x/y
                    reads in M13mp19, .b/g reads in pUC19). The same ligation
                    was transformed into DH5a."

ORIGIN
Alignment Scores:      Length:      724
Pred. No.:            58.7          Matches:      34
Score:                76.00         Conservative: 7
Percent Similarity:   36.28%        Mismatches:   29
Best Local Similarity: 30.09%        Indels:       43
Query Match:          16.20%        Gaps:         6
DB:                   8

US-09-917-376-5 (1-88) x BZ616323 (1-724)

QY      2 SerGlyVallysValGlnTyrLysAsnAsnAspSer-----AlaPro 16
      :::::|||||:::|||||:::
DB      587 TCGCGAGCGTCAGCATCCCATCAAGAGCGGCAACGACGCTGGCGAGTCGGGCC 528

QY      17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
      :::::|||||:::|||||:::
DB      527 GGGCGCCAGCTGAGCAGCTGCAGCGGAGCGACGACTGATGCTTATCAACCGTGGCGGC 468

QY      30 Thr-GlySerSerSerValAsp-----
      :::::|||||:::|||||:::
DB      467 ACGAGGAAGTTCTGTCACACTTCCCGGAGATCGTCGGAGCGGTCCAGAGCGTGGGTTCGAG 408

QY      37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 53
      :::::|||||:::|||||:::
DB      407 GTCATCCCGATCGAGCGCGTAGAGACTCAGCGGTGAGGAGCTTCGCCGCGACCGCTCGAC 348

QY      53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
      :::::|||||:::|||||:::
DB      347 TCGTGGC-----ACGTGCTCATGGGCGCGCACGGG----- 317
```



```
Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 316 -----CGGCGCTCACCACTTCTCTTCC 293

RESULT 12
B1218457 814 bp mRNA linear EST 11-JUL-2001
LOCUS 602937951P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5101069 5',
DEFINITION mRNA sequence.
ACCESSION B1218457
VERSION B1218457.1 GI:14671901
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 814)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11243 row: f column: 14
High quality sequence stop: 792.
FEATURES
source
1..814
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5101069"
/lab_host="DH10B (TI phage-resistant)"
/clone_lib="NCI_CGAP_L19"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
ORIGIN
Alignment Scores:
Pred. No.: 68.3 Length: 814
Score: 76.00 Matches: 28
Percent Similarity: 44.83% Conservative: 11
Best Local Similarity: 32.18% Mismatches: 29
Query Match: 16.20% Indels: 19
DB: 4 Gaps: 4
US-09-917-376-5 (1-88) x B1218457 (1-814)
Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 33 GCAGGTGAGCCCGAGTATGGAGCGGAACACACAGATCGCCAGGCAACATGAGCAG-- 89
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 90 -----AGTTCCCGAGTGGAGCCTGAGGACCCCTG--- 116
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeu-----ValTyrAsnCys 59
Db 117 CTCGCTACTACACACAGCGCGCGCTCTACAGATCCAGCGGATGACGCTGT 176
Qy 60 Asp-TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAl 79
Db 177 GACGTGGGTGCGACGGCGCC-----TCCTCCGCGGTACGAGCAGGTGCGCA 224
Qy 79 aThrProThrAlaAspThr 85

RESULT 13
B2545807 905 bp DNA linear GSS 16-DEC-2002
LOCUS OGALR72TM ZM2 0.7 1.5 KB Zea mays genomic clone ZMMBma0089L24,
DEFINITION genomic survey sequence.
ACCESSION B2545807
VERSION B2545807.1 GI:27099699
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 905)
REFERENCE Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
AUTHORS Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
JOURNAL Consortium for Maize Genomics
COMMENT Unpublished (2002)
Other GSSs: OGALR72TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..905
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBma0089L24"
/clone_lib="ZM2 0.7 1.5 KB"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN
Alignment Scores:
Pred. No.: 78.3 Length: 905
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.20% Indels: 43
DB: 8 Gaps: 6
US-09-917-376-5 (1-88) x B2545807 (1-905)
Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 87 TCCGCGCGCTGAGCATCCATACAGAGGCGGCAACGCTCGGCGAGTGGCCC 146
Qy 17 Gly-----AspAsnGlnIleLysProGlyLeuGlnLeuValAsn----- 29
Db 147 GGCGCCCGCAGCTGAGCAGCTGCAGCGGAGGCCACGACTGATGCTTATCAACCGTGGCGC 206
Qy 30 Thr-GlySerSerValAsp----- 36
Db 207 ACGAGGAAGTTCTGCTCAACTTCCCGAGATCGTCGAGCGGTCCAGAGGCTGGGTTTCGAG 266
Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlySerSe 53
Db 267 GTCATCCGATCGAGCCGCTAGAGACCTCAGCGTGAGGACTTCGCCCGGAGCCGTGAC 326
Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db 327 TCGTGG-----ACGTGCTCATGGCGCGCAGGG----- 357
```



```
Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 358 -----CGGCGCTCACCACTTCTTCTTCC 381

RESULT 14
CG270741/c
LOCUS CG270741 934 bp DNA linear GSS 25-AUG-2003
DEFINITION OGWIB33TH ZM 0.7_1.5 KB Zea mays genomic clone ZMMBMA0583E18,
genomic survey sequence.
ACCESSION CG270741
VERSION CG270741.1 GI:34182882
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 934)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
TITLE Unpublished (2002)
JOURNAL Other GSSs: OGWIB33TH
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..934
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0583E18"
/clone_lib="ZM 0.7_1.5 KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN
Alignment Scores:
Pred. No.: 81.5 Length: 934
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.20% Indels: 43
DB: 9 Gaps: 6
US-09-917-376-5 (1-88) x CG270741 (1-934)
Qy 2 SerGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 804 TCCGAGCGTCAGCATCCATACAAAGAGGCGGACACGACGCTGCGCAGCTGGCGCC 745
Qy 17 Gly-----AspAsnGlnLysProGlyLeuGlnLeuValAsn----- 29
Db 744 GCGGCCCGAGCTGAGCAGCTGACGCGGAGGAGCCACGACTGATGCTTATCAACCGTGGCCGC 685
Qy 30 Thr-GlySerSerValAsp----- 36
Db 684 ACAGAGAGTTCTGCTCACTTCCGAGAGATCGTCGAGCGGTCCAGAGGCTGGGTTCGAG 625
Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db 624 GTCATCCGATGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 565
Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db 564 TCGTGG-----ACGTGCTCATGGCGCGCGCGGG----- 534

Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 533 -----CGGCGCTCACCACTTCTTCTTCC 510

RESULT 15
CG270748
LOCUS CG270748 1001 bp DNA linear GSS 25-AUG-2003
DEFINITION OGWIB33TV ZM 0.7_1.5 KB Zea mays genomic clone ZMMBMA0583E18,
genomic survey sequence.
ACCESSION CG270748
VERSION CG270748.1 GI:34182889
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
TITLE Unpublished (2002)
JOURNAL Other GSSs: OGWIB33TH
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..1001
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0583E18"
/clone_lib="ZM 0.7_1.5 KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN
Alignment Scores:
Pred. No.: 89.1 Length: 1001
Score: 76.00 Matches: 34
Percent Similarity: 36.28% Conservative: 7
Best Local Similarity: 30.09% Mismatches: 29
Query Match: 16.20% Indels: 43
DB: 9 Gaps: 6
US-09-917-376-5 (1-88) x CG270748 (1-1001)
Qy 2 SerGlyValLysValGlnTyrLysAsnAspSer-----AlaPro 16
Db 418 TCCGAGCGCTCAGCATCCATACAAAGAGGCGGACACGCTGCGCAGCTGGCGCC 477
Qy 17 Gly-----AspAsnGlnLysProGlyLeuGlnLeuValAsn----- 29
Db 478 GCGGCCCGAGCTGAGCAGCTGACGCGGAGGAGCCACGACTGATGCTTATCAACCGTGGCCGC 537
Qy 30 Thr-GlySerSerValAsp----- 36
Db 538 ACAGAGAGTTCTGCTCACTTCCGAGAGATCGTCGAGCGGTCCAGAGGCTGGGTTCGAG 597
Qy 37 -----LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSe 53
Db 598 GTCATCCGATGAGCGCGTAGAGACCTCAGCGTGGAGGACTTCGCCCGGACCGTCGAC 657
Qy 53 rThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPh 73
Db 533 -----ACGTGCTCATGGCGCGCGCGGG----- 534
```


Db 658 TCGTGGC-----ACGTGCTCATGGCGCGCACGGGG----- 688
Qy 73 eGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 689 -----CCGGCCTCACCAACTTCTTCTCC 712

Search completed: October 5, 2005, 21:09:34
Job time : 848.19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 07:47:30 ; Search time 41.8869 Seconds
(without alignments)
3437.649 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKQYKXNDSPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -US09917376/runat 04102005 164328 26925/app query.fasta_l.2524
-Q=/cgn2_1/USPTO.spool/US09917376 -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-DB=Issued Patents NA -QWTF=fastcap -SUFFIX=blosum62 -TRANS=human40.cdi
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09917376 @cgn 1.1 516 @runat 04102005 164328 26925 -NCFU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	41.7	2029	3	US-09-136-574A-46 Sequence 46, Appl
2	194.5	41.5	6416	3	US-09-136-574A-2 Sequence 2, Appl
3	192.5	41.0	11707	3	US-09-136-574A-1 Sequence 1, Appl
4	182.5	38.9	2977	2	US-07-862-588B-1 Sequence 1, Appl
5	165	35.2	1438	4	US-09-339-159B-3 Sequence 3, Appl
6	165	35.2	1482	3	US-09-198-956-9 Sequence 9, Appl
7	165	35.2	1482	3	US-09-198-955A-11 Sequence 11, Appl
8	165	35.2	1482	3	US-09-694-531-11 Sequence 11, Appl
9	165	35.2	1482	3	US-09-670-141-9 Sequence 9, Appl
10	165	35.2	1482	4	US-10-072-152-11 Sequence 11, Appl
11	155.5	33.2	4059	4	US-09-784-554B-1 Sequence 1, Appl
12	150.5	32.1	4056	4	US-09-784-554B-3 Sequence 3, Appl

13	120.5	25.7	1624	2	US-07-862-588B-6 Sequence 6, Appl
14	120.5	25.7	1775	2	US-07-862-588B-5 Sequence 5, Appl
15	114	24.3	486	1	US-08-048-164A-2 Sequence 2, Appl
16	114	24.3	486	1	US-08-048-164A-3 Sequence 3, Appl
17	114	24.3	486	1	US-08-460-462-1 Sequence 1, Appl
18	114	24.3	486	1	US-08-460-462-3 Sequence 3, Appl
19	114	24.3	486	1	US-08-460-457-1 Sequence 1, Appl
20	114	24.3	486	1	US-08-460-457-3 Sequence 3, Appl
21	114	24.3	486	1	US-08-460-458-1 Sequence 1, Appl
22	114	24.3	486	1	US-08-460-458-3 Sequence 3, Appl
23	114	24.3	486	2	US-08-460-455-1 Sequence 1, Appl
24	114	24.3	486	2	US-08-460-455-3 Sequence 3, Appl
25	114	24.3	486	2	US-08-330-394A-1 Sequence 1, Appl
26	114	24.3	486	2	US-08-330-394A-3 Sequence 3, Appl
27	114	24.3	499	3	US-09-006-636-5 Sequence 5, Appl
28	114	24.3	499	3	US-09-006-632-5 Sequence 5, Appl
29	114	24.3	499	3	US-09-325-274-5 Sequence 5, Appl
30	113	24.1	1146	3	US-09-277-716-21 Sequence 21, Appl
31	113	24.1	1146	3	US-09-609-161B-21 Sequence 21, Appl
32	77	16.4	8107	4	US-09-335-586-3 Sequence 3, Appl
33	76	16.2	136058	4	US-09-949-016-12565 Sequence 12565, A
34	76	16.2	136480	4	US-09-949-016-17064 Sequence 17064, A
35	75.5	16.1	5714	4	US-09-620-312D-393 Sequence 393, App
36	72.5	15.5	819	4	US-09-902-540-3752 Sequence 3752, Ap
37	72.5	15.5	17897	4	US-09-902-540-1182 Sequence 1182, Ap
38	72.5	15.5	20284	4	US-09-526-193A-21 Sequence 21, Appl
39	70	14.9	10892	4	US-09-902-540-962 Sequence 962, App
40	70	14.9	25165	3	US-09-453-702B-39 Sequence 39, Appl
41	70	14.9	246444	4	US-09-949-016-13113 Sequence 13113, A
42	69.5	14.8	103987	4	US-09-949-016-12513 Sequence 12513, A
43	69.5	14.8	103988	4	US-09-949-016-17050 Sequence 17050, A
44	68.5	14.6	601	4	US-09-949-016-128395 Sequence 128395,
45	68.5	14.6	601	4	US-09-949-016-128396 Sequence 128396,

ALIGNMENTS

RESULT 1
US-09-136-574A-46
; Sequence 46, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbons, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSER: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136.574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 2029 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-136-574A-46

Alignment Scores:
Pred. No.: 6.46e-17 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.23% Conservative: 15
Best Local Similarity: 43.18% Mismatches: 34
Query Match: 41.68% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-136-574A-46 (1-2029)

Qy 1 ValSerGlyValLeuGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1 ATGGGAAGTGGTGAAGGACTGTACAGAACAATGAGCAAGTGGCAGCAGGTTCT 60
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 61 ATAAGGCGTGGTTAAGATAGTGAATGGAGCAGCAGCAGTGTGTATCTTAGCAGGTT 120
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 121 AAGATAGACTACTGGTACACAGTGGATGGTGTGACAAAGCCACAGAGTGGCGGTA 177
Qy 61 TrpAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTG 237
Qy 81 ProThrAlaAspThrTyrLeuGln 88
Db 238 AGTGAGCGGATTTACCTGGAG 261

RESULT 2

US-09-136-574A-2
Sequence 2, Application US/09136574A
Patent No. 6294366

GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6416 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2

Alignment Scores:
Pred. No.: 4.68e-16 Length: 6416
Score: 194.50 Matches: 38
Percent Similarity: 61.18% Conservative: 14
Best Local Similarity: 44.71% Mismatches: 32
Query Match: 41.47% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-136-574A-2 (1-6416)

Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2665 GGTGTGAAGTACTGTACAGAACAATGAGCAAGTGGCAGCAGGTTCTTATAGGCGG 2724
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 2725 TGGTTTAAGTAGTGAATGGAGCGCAGCAGCAGTGTGTATCTTAGCAGGGTTAAGATAAGA 2784
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 2785 TACTGTACACAGTGGTGGTGCACAAAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 2841
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2842 ATAGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTGTAGTGGAGCG 2901
Qy 84 AspThrTyrLeuGln 88
Db 2902 GATTATTACCTGGAG 2916

RESULT 3

US-09-136-574A-1
Sequence 1, Application US/09136574A
Patent No. 6294366

GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House


```
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/136,574A
  FILING DATE: 19-Aug-1998
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/932,571
  FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
  NAME: Bak, Mary E.
  REGISTRATION NUMBER: 31,215
  REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 215-540-9200
  TELEFAX: 215-540-5818
  TELEX: <Unknown>
SEQUENCE CHARACTERISTICS:
  LENGTH: 11707 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-136-574A-1

Alignment Scores:
Pred. No.: 2,11e-15 Length: 11707
Score: 192.50 Matches: 37
Percent Similarity: 62.3% Conservative: 16
Best Local Similarity: 43.5% Mismatches: 31
Query Match: 41.0% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-136-574A-1 (1-11707)
QY 4 GlyValLeuValGlnTyrIleAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 4038 GGTTCGAGGTACTATACAGAACATGACAGCGGACGACAGGCTCTATAGCGCG 4097
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 4098 TGTGTTAAGATAGTGAATGAGGACGACAGGCTGTGATCTTAGCGGGTTAAGATA 4157
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 4158 TACTGTTACACAGTGTGTGACAGCCACAGAGTGGGTA---TGTGACTGGGACAG 4214
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 4215 ATAGGCGCAAGCAATGTCATTCATTTTGTGAGCTGACGAGCGGAGTGAGTGAGCG 4274
QY 84 AspThrTyrLeuGln 88
Db 4275 GATTATTACTTGGAG 4289

RESULT 4
US-09-917-376-5 (1-88) x US-09-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnae
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.

STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/07/862,588B
  FILING DATE: 19920727
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: DK 164/90
    FILING DATE: 19-JAN-1990
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/DK91/00013
    FILING DATE: 18-JAN-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Zelson, Steve T. / Lambiris, Elias J.
    REGISTRATION NUMBER: 30,335 / 33,728
    REFERENCE/DOCKET NUMBER: 3425.204-US
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 212 867 0123
    TELEFAX: 212 867 0298
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 2977 base pairs
      TYPE: NUCLEIC ACID
      STRANDEDNESS: double
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    ORIGINAL SOURCE:
      ORGANISM: Bacillus lautus
      STRAIN: NCIMB 40250
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 677..2776
      OTHER INFORMATION:
        US-07-862-588B-1

Alignment Scores:
Pred. No.: 7,14e-15 Length: 2977
Score: 182.50 Matches: 36
Percent Similarity: 66.27% Conservative: 19
Best Local Similarity: 43.37% Mismatches: 25
Query Match: 38.91% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-1 (1-2977)
QY 6 LysValGlnTyrIleAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25
Db 144 CGACTCGATTACAGCGCGCGGATACAAATCCAGCCGACACACAGATCAAGCCGCTTC 203
QY 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45
Db 204 AACATCAAAACAAACGCTACTTCGGCTGTGATTAAAGCAGCGCTCAAAATCCGCTACTAC 263
QY 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
Db 264 TTCACCAAGATGTTCTTCGCGCGGTGACAGCGTGG---ATCGACTGGGCGCAGCTCGGC 320
QY 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 321 GGCAGCAACATTCAGATCTCTGTTTGGC-----AACCATCTGGCAGCAATTCGATACG 374
QY 86 TyrLeuGln 88
Db 375 TACGTGGAG 383
```



```
RESULT 5
US-09-339-159B-3
; Sequence 3, Application US/09339159B
; Patent No. 6566114
; ORGANISM: Bacillus licheniformis
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339.159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. 1633
US-09-339-159B-3

Alignment Scores:
Pred. No.: 6.68e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 4 Gaps: 2

US-09-917-376-5 (1-88) x US-09-339-159B-3 (1-1438)
Qy 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTGAGGTTGAATCTTACACGAGCATCTTCAGATCTACTAACTCA 999
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCTCTCAGTTCAGTTACTAATACCGAAGCATGCAATTTGTCCAAATCTC 1059
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTACGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1116
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATAATCGCGAGTAAACGCGAGCTACAAACGGAATTACTTCAAAATGTAAAGGA 1176
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAATAATGAGTTCTTCAACAATAATACGACAGACCTACTCTTGAA 1227

RESULT 6
US-09-198-956-9
; Sequence 9, Application US/09198956
; Patent No. 6165769
; ORGANISM: Bacillus licheniformis
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198.956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-198-956-11

; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-198-956-9

Alignment Scores:
Pred. No.: 6.97e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-198-956-9 (1-1482)
Qy 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGAGGTTGAATCTTACACGAGCATCTTCAGATCTACTAACTCA 1050
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCTCTCAGTTCAGTTACTAATACCGAAGCATGCAATTTGATTTGTCCAAATCTC 1110
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGCGAGTAAACGCGAGCTACAAACGGAATTACTTCAAAATGTAAAGGA 1227
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAATAATGAGTTCTTCAACAATAATACGACAGACCTACTCTTGAA 1278

RESULT 7
US-09-198-955A-11
; Sequence 11, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Samne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198.955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 05/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-09-198-955A-11
```


Alignment Scores:		6.97e-13		1482	
Pred. No.:	Score:	165.00	Matches:	35	
Percent Similarity:	59.79%	Conservative:	23		
Best Local Similarity:	36.08%	Mismatches:	29		
Query Match:	35.18%	Indels:	10		
DB:	3	Gaps:	2		
US-09-917-376-5 (1-88) x US-09-198-955A-11 (1-1482)					
Qy	1	ValSerGlyGlyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln	20		
Db	991	GTATCAGCAATTTGAGGTTGAATCTACACAGCAATCTTCAGATACTACTCA	1050		
Qy	21	lIeLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal	40		
Db	1051	ATCAATCCTCAGTTCAGGTTACTAATACCGAGCAGTGCAATTTGTCCTCAAACTC	1110		
Qy	41	ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp	60		
Db	1111	ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC	1167		
Qy	61	TpAlaAlaMet-----GlyCysGlyAsnIleArgAla	71		
Db	1168	CATGCTGCAATATCGGCAGTAGACGAGCTACACGGAATTACTTCAATGTAAAGGA	1227		
Qy	72	SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln	88		
Db	1228	ACATTGTGTAATAATGAGTTCCTCAACAATAACGACACCTACCTTGAA	1278		
RESULT 8					
US-09-694-531-11					
; Sequence 11, Application US/09694531					
; Patent No. 6368843					
; GENERAL INFORMATION:					
; APPLICANT: Andersen, Lene N.					
; APPLICANT: Schulein, Martin					
; APPLICANT: Lange, Niels Erik E.					
; APPLICANT: Moller, Soren					
; APPLICANT: Glad, Sanne O. S.					
; APPLICANT: Kauppinen, Markus S.					
; APPLICANT: Kongsbak, Lars					
; TITLE OF INVENTION: No. 6368843el Pectate Lyases					
; FILE REFERENCE: 5378 200-US					
; CURRENT APPLICATION NUMBER: US/09/694,531					
; PRIOR FILING DATE: 2000-10-23					
; PRIOR APPLICATION NUMBER: 09/198,955					
; PRIOR FILING DATE: 1998-11-24					
; PRIOR APPLICATION NUMBER: 1343/97					
; PRIOR FILING DATE: 1997-11-24					
; PRIOR APPLICATION NUMBER: 1344/97					
; PRIOR FILING DATE: 1997-11-24					
; PRIOR APPLICATION NUMBER: 60/067,249					
; PRIOR FILING DATE: 1997-12-02					
; PRIOR APPLICATION NUMBER: 60/067,240					
; PRIOR FILING DATE: 1997-12-02					
; PRIOR APPLICATION NUMBER: 09/073,684					
; PRIOR FILING DATE: 1998-05-06					
; PRIOR APPLICATION NUMBER: 09/184,217					
; PRIOR FILING DATE: 1998-11-02					
; NUMBER OF SEQ ID NOS: 32					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 11					
; LENGTH: 1482					
; TYPE: DNA					
; ORGANISM: Clostridium thermocellum					
US-09-694-531-11					
Alignment Scores:		6.97e-13		1482	
Pred. No.:	Score:	165.00	Matches:	35	
Percent Similarity:	59.79%	Conservative:	23		
Best Local Similarity:	36.08%	Mismatches:	29		
Query Match:	35.18%	Indels:	10		
DB:	3	Gaps:	2		

Alignment Scores:		6.97e-13		1482	
Pred. No.:	Score:	165.00	Matches:	35	
Percent Similarity:	59.79%	Conservative:	23		
Best Local Similarity:	36.08%	Mismatches:	29		
Query Match:	35.18%	Indels:	10		
DB:	3	Gaps:	2		
US-09-917-376-5 (1-88) x US-09-694-531-11 (1-1482)					
Qy	1	ValSerGlyGlyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln	20		
Db	991	GTATCAGCAATTTGAGGTTGAATCTACACAGCAATCTTCAGATACTACTCA	1050		
Qy	21	lIeLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal	40		
Db	1051	ATCAATCCTCAGTTCAGGTTACTAATACCGAGCAGTGCAATTTGTCCTCAAACTC	1110		
Qy	41	ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp	60		
Db	1111	ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC	1167		
Qy	61	TpAlaAlaMet-----GlyCysGlyAsnIleArgAla	71		
Db	1168	CATGCTGCAATATCGGCAGTAGACGAGCTACACGGAATTACTTCAATGTAAAGGA	1227		
Qy	72	SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln	88		
Db	1228	ACATTGTGTAATAATGAGTTCCTCAACAATAACGACACCTACCTTGAA	1278		
RESULT 9					
US-09-670-141-9					
; Sequence 9, Application US/09670141					
; Patent No. 6429000					
; GENERAL INFORMATION:					
; APPLICANT: Andersen, Lene N.					
; APPLICANT: Schulein, Martin					
; APPLICANT: Lange, Niels Erik K.					
; APPLICANT: Bjornvad, Mads E.					
; APPLICANT: Schnorr, Kirk					
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus					
; FILE REFERENCE: 5377.200-US					
; CURRENT APPLICATION NUMBER: US/09/670,141					
; CURRENT FILING DATE: 2000-09-26					
; PRIOR APPLICATION NUMBER: 09/198,956					
; PRIOR FILING DATE: 1998-11-24					
; PRIOR APPLICATION NUMBER: 1344/97					
; PRIOR FILING DATE: 1997-11-24					
; PRIOR APPLICATION NUMBER: 60/067,240					
; PRIOR FILING DATE: 1997-12-02					
; NUMBER OF SEQ ID NOS: 26					
; SOFTWARE: FastSeq for Windows Version 3.0					
; SEQ ID NO 9					
; LENGTH: 1482					
; TYPE: DNA					
; ORGANISM: Bacillus licheniformis					
US-09-670-141-9					
Alignment Scores:		6.97e-13		1482	
Pred. No.:	Score:	165.00	Matches:	35	
Percent Similarity:	59.79%	Conservative:	23		
Best Local Similarity:	36.08%	Mismatches:	29		
Query Match:	35.18%	Indels:	10		
DB:	3	Gaps:	2		
US-09-917-376-5 (1-88) x US-09-670-141-9 (1-1482)					
Qy	1	ValSerGlyGlyValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGln	20		
Db	991	GTATCAGCAATTTGAGGTTGAATCTACACAGCAATCTTCAGATACTACTCA	1050		
Qy	21	lIeLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal	40		

[illegible]

```

Db      1111 ACATTGAGATATTATTATACAGTAGACGGCAGAAAGATCAGACCTTCTCG---TGTCAG 1167
Qy      61   TrpAlaAalaMet-----GlyCysGlyAsnIleArgAla 71
Db      1168 CATGCTGCATAATCCGCCAGTAACAACGGAAATTACTCAATGTAAAAAGGA 1227
Qy      72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrlLeuGln 88
Db      1228 ACATTGTAAAATGAGTTCTCTCAACAAAATAACGCAGACACCTTACCTTGAA 1278

RESULT 11
US-09-784-554B-1
; Sequence 1, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784, 554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4059
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-1

Alignment Scores:
Pred. No.:          6.2e-11           Length:         4059
Score:             155.50            Matches:        27
Percent Similarity: 56.47%           Conservative:   21
Best Local Similarity: 31.76%         Mismatches:    36
Query Match:       33.16%            Indels:         1
DB:                  4                Gaps:           1

US-09-917-376-5 (1-88) x US-09-784-554B-1 (1-4059)
Qy      4   GlyValysValGlnTrtylsAsnAsnAspSerAlaProGlyAspGlnIleLysPro 23
Db      3598 GGAGCTGCTGCTCAATATCCGCACAGCAGATACTAAGTGTAACGATAATCACCTCAATPCG 3657
Qy      24   GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db      3658 CAATTCAAATTTGTAACAAAGGCACAACTCCATACCGATCAACGAGTTGAAAATTCGC 3717
Qy      44   TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyraenCysAspTrpAlaAala 63
Db      3718 TACTACTACACAATCGACCGGTGACCGGTGAGCAG--ACTTTTCAACTGCGACTATGCGACG 3774
Qy      64   MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db      3775 CTGAGCTGCTCAAAGCTGAACGTAACCTGTTTAAAATGAGAAGGCTGCAACCGGTGCC 3834
Qy      84   AspThrTyrlLeuGln 88
Db      3835 GATTATTATTGAA 3849

RESULT 12
US-09-784-554B-3
; Sequence 3, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784, 554B
; CURRENT FILING DATE: 2001-02-16
```



```
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 3,06e-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.4% Conservative: 21
Best Local Similarity: 31.7% Mismatches: 36
Query Match: 32.0% Indels: 1
DB: 4 Gaps: 1

US-09-917-376-5 (1-88) x US-09-784-554B-3 (1-4056)
QY 4 GlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3595 GGAATGGTGCTCCGATATCGCACGCGGATACAAATGTGAACACATCACTTGACCCG 3654

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3655 CATTTCCTCAATTTTAATAAAGGTACAAATCTCCGTACCGATCAACGAGTTGAAATTCG 3714

QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63
Db 3715 TACTACTACAGCATCGCGTGACCGTGACGAG---ACATTCAACTGCGACTATGCGGTG 3771

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThra 83
Db 3772 CTGAGCTGCTCGAAGCTGAATCGTAAGCTGGTTAAATATGATAAAGCTGCAACCGGTGCT 3831

QY 84 AspThrTyrLeuGln 88
Db 3832 GATTATTATTGGAA 3846

RESULT 13
US-07-862-588B-6
; Sequence 6, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728

; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1607
; OTHER INFORMATION:
US-07-862-588B-6

Alignment Scores:
Pred. No.: 1.18e-06 Length: 1624
Score: 120.50 Matches: 28
Percent Similarity: 54.12% Conservative: 18
Best Local Similarity: 32.94% Mismatches: 36
Query Match: 25.69% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-6 (1-1624)
QY 2 SerGlyGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 1323 ACGGGGAACCTTTGTCCTCAATACAAAGTTGGGACACTAGCCGACGATAACCAATG 1382

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 1383 AGCCCTTCCTTTAAACATCAAGAACACGGGTACAAACCCCTGTTAACTGAGCGGCTCAAG 1442

QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 1443 CTTNNNNNNNNNNNNNNAAAAAGAC---GGACCTGCGGATATGAGCTGCTCGATCGACTGG 1499

QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 1500 GCGCAAAATCGCCGACGAAGATGTTCTGCTGCGCATTC-----GCTAACTTACCGGGAGT 1553

QY 82 ThrAlaAspThrTyr 86
Db 1554 ATACCGATACTTAC 1568

RESULT 14
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
```



```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1775 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..(1625..1775)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /partial
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
US-07-862-588B-5

Alignment Scores:
Pred. No.: 1.34e-06 Length: 1775
Score: 120.50 Matches: 28
Percent Similarity: 54.12% Conservative: 18
Best Local Similarity: 32.94% Mismatches: 36
Query Match: 25.69% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-5 (1-1775)
QY 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 1323 ACGGGGAACCTTGTGTCCATACAAAGTTGCGGACCTAGCGCCACGGATAACCAATG 1382
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 1393 AAGCTTCTTAACTACATCAAGACACGCTAGACACCTCTGTAACTGAGCGGCTCAAG 1442
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 1443 CTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1499
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 1500 CGCAAAATCGCGCGCAAGCAAGTGTCTGCTGGCATTC-----GCTAACTTTACCGGGAGT 1553
QY 82 ThrAlaAspThrTyr 86
DB 1554 AATCGGATACCTTAC 1568

RESULT 15
US-08-048-164A-1
; Sequence 1, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
US-08-048-164A-1

Alignment Scores:
Pred. No.: 1.65e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.58% Conservative: 22
Best Local Similarity: 29.90% Mismatches: 34
Query Match: 24.31% Indels: 12
DB: 1 Gaps: 4

US-09-917-376-5 (1-88) x US-08-048-164A-1 (1-486)
QY 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 7 ACATCATCAATGCTGAGTTGAATTTTCAACTCTAACAAATCAGCACAAACAACTCAATT 66
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 67 ACACCAATAAATAAAATTAATAACATCTGACATCTGATTTAAATTTAAATGACGTAAAA 126
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
DB 127 GTTAGATATATTATACACAGTGTGTTACACAGGACAACTTTCTTGG---TGTGACCAT 183
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
DB 184 GCTGTGCTATATTAGGAATAGTATGTTGATACACTAGCAAGTACGACGCAACTTC 243
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
DB 244 GTTAAAGAAACAGCAAGCCCA---ACATCACTATGATACATATGTTGAA 291
```


Search completed: October 5, 2005, 22:57:00
Job time : 47.8869 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 08:30:02 ; Search time 169.52 Seconds
(without alignments)
3584.301 Million cell updates/sec

Title: US-09-917-376-5
Perfect score: 469
Sequence: 1 VSGGVKQYKKNDSAPGDNQ.....IRASFGSVNPTATDTYLQ 88

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09917376/runat_04102005_164329_26954/app_query.fasta_1.2524
-DB=Published Applications NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09917376 -CGCN 1 1 2178 @runat_04102005_164329_26954
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCX=100
-LOGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469	100.0	2289	10	US-09-917-378-2 Sequence 2, Appli
2	466	99.4	3687	10	US-09-917-384-2 Sequence 2, Appli
3	466	99.4	3687	10	US-09-917-383-2 Sequence 2, Appli
4	462	98.5	2869	11	US-09-917-376-2 Sequence 2, Appli
5	462	98.5	2869	15	US-10-155-400-2 Sequence 2, Appli
6	257	54.8	2223	15	US-10-156-761-550 Sequence 550, App
7	257	54.8	9025608	15	US-10-156-761-1 Sequence 1, Appli
8	190	40.5	1527	17	US-10-369-493-46838 Sequence 46838, A
9	188	40.1	1434	19	US-10-466-208-11 Sequence 11, Appl
10	188	40.1	1491	19	US-10-466-208-7 Sequence 7, Appli
11	188	40.1	2510	19	US-10-433-577-34 Sequence 34, Appl
12	165	35.2	501	21	US-10-933-404-1 Sequence 1, Appli
13	165	35.2	1438	17	US-10-372-054-3 Sequence 3, Appli
14	165	35.2	1482	13	US-10-072-152-11 Sequence 11, Appl
15	165	35.2	1482	18	US-10-655-433-11 Sequence 11, Appl
16	157	33.5	1314	9	US-09-974-300-684 Sequence 684, App
17	155.5	33.2	4059	10	US-09-784-554B-1 Sequence 1, Appli
18	155.5	33.2	4059	20	US-10-896-555-1 Sequence 1, Appli
19	150.5	32.1	4056	10	US-09-784-554B-3 Sequence 3, Appli
20	150.5	32.1	4056	20	US-10-896-555-3 Sequence 3, Appli
21	131.5	28.0	3180	22	US-10-886-393A-29 Sequence 29, Appl
22	130.5	27.8	4452	17	US-10-282-122A-15299 Sequence 15299, A
23	126.5	27.0	1303	21	US-10-490-737-1 Sequence 1, Appli
24	126.5	27.0	1747	21	US-10-490-737-2 Sequence 2, Appli
25	126.5	27.0	3489	21	US-10-490-737-3 Sequence 3, Appli
26	114	24.3	768	16	US-10-261-446-5 Sequence 5, Appli
27	114	24.3	768	18	US-10-261-448B-5 Sequence 5, Appli
28	114	24.3	768	21	US-10-782-234-5 Sequence 5, Appli
29	113	24.1	652	20	US-10-748-055-9 Sequence 9, Appli
30	113	24.1	1146	10	US-09-808-898-21 Sequence 21, Appl
31	78.5	16.7	764	20	US-10-653-047-5404 Sequence 5404, Ap
32	78.5	16.7	3147	17	US-10-282-122A-13716 Sequence 13716, A
33	76	16.2	96597	18	US-10-052-482-226 Sequence 226, App
34	76	16.2	757	19	US-10-767-701-8547 Sequence 742, App
35	75.5	16.1	1041	17	US-10-260-238-742 Sequence 742, App
36	75.5	16.1	1041	19	US-10-437-963-76751 Sequence 76751, A
37	75.5	16.1	2091	18	US-10-211-462-222 Sequence 222, App
38	75.5	16.1	2091	20	US-10-723-860-4038 Sequence 4038, Ap
39	75.5	16.1	2091	22	US-10-756-149-4006 Sequence 4006, Ap
40	75.5	16.1	3396	17	US-10-398-458-1 Sequence 1, Appli
41	75.5	16.1	4041	9	US-09-747-835A-21 Sequence 21, Appl
42	75.5	16.1	4041	18	US-10-312-312-21 Sequence 21, Appl
43	75.5	16.1	4353	15	US-10-225-567A-405 Sequence 405, App
44	75.5	16.1	4755	21	US-10-505-486-201 Sequence 201, App
45	75.5	16.1	4914	9	US-09-974-298-188 Sequence 188, App

ALIGNMENTS

RESULT 1
US-09-917-378-2
; Sequence 2, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; FILE REFERENCE: 40197.7US01
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8

Pred. No.: 8.85e-26 Length: 2223
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.80% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-5 (1-88) x US-10-156-761-550 (1-2223)

Qy 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 1771 TCGGCGCGCTTCCAGGTCCTCTCAAGAACACGACTCTCGGCCACCGCAACGCCCATC 1830
Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 1831 CGGCGCGCTTCCAGGTCCTCTCAAGAACACGACTCTCGGCCACCGCAACGCCCATC 1890
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 1891 GCCCGCTACTTCCACCGGACAGCGCTCGGCCACCGTGAACGCTGTCGCGACTAC 1950
Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 1951 GCGCGCGCTTCCAGGTCCTCTCAAGAACACGACTCTCGGCCACCGCAACGCCCATC 2010
Qy 82 ThrAlaAspThrTyrLeuGln 88
Db 2011 GGAGCGGAGCTTACTCGAA 2031

RESULT 7

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Alignment Scores:
Pred. No.: 5.62e-21 Length: 9025608
Score: 257.00 Matches: 49
Percent Similarity: 71.26% Conservative: 13
Best Local Similarity: 56.32% Mismatches: 25
Query Match: 54.80% Indels: 0
DB: 15 Gaps: 0

US-09-917-376-5 (1-88) x US-10-156-761-1 (1-9025608)

Qy 2 SerGlyGlyValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 706216 TCGGCGCGCTTCCAGGTCCTCTCAAGAACACGACTCTCGGCCACCGCAACGCCCATC 706275

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 706276 CGGCGCGCTTCCAGGTCCTCTCAAGAACACGACTCTCGGCCACCGCAACGCCCATC 706335
Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 706336 GCGCGCTTACTTCCACCGGACAGCGCTCGGCCACCGTGAACGCCCTGTCGCGACTAC 706395
Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 706396 GCGGCGGTCGCTTCCAGGTCCTCTCAAGAACACGACTCTCGGCCACCGCAACGCCCATC 706455
Qy 82 ThrAlaAspThrTyrLeuGln 88
Db 706456 GGAGCGGAGCTTACTCGAA 706476

RESULT 8

US-10-369-493-46838
; Sequence 46838, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46838
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-369-493-46838

Alignment Scores:

Pred. No.: 1.94e-16 Length: 1527
Score: 190.00 Matches: 35
Percent Similarity: 65.12% Conservative: 21
Best Local Similarity: 40.70% Mismatches: 26
Query Match: 40.51% Indels: 4
DB: 17 Gaps: 2

US-09-917-376-5 (1-88) x US-10-369-493-46838 (1-1527)

Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 1084 GGTATTCTGTACAGTACAGCAGCGAGGATGATGAACAGCAACCAATCCGTCG 1143
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1144 CAGCTTCAATAAATAAATAACCGCAATACCGGTTGATTAAAGATGTCATGCCCGT 1203
Qy 44 TyrTrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62
Db 1204 TACTGTATTAAGCGAATAAACAAGGCCCAAC-----TTTGACTGTGACTAGCG 1254
Qy 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82
Db 1255 CAGATTGGATGCGCAATGTGACACACAAAGTTTGTGCGTTGTCATAAACAAGCAAGGT 1314
Qy 83 AlaAspThrTyrLeuGln 88
Db 1315 GCAGATACCTATCTGAA 1332

RESULT 9

US-10-466-208-11


```
; Sequence 11, Application US/10466208
; Publication No. US20040180348A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 11
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCase gene with additional his encoding sequences
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1431)
US-10-466-208-11

Alignment Scores:
Pred. No.: 3,44e-16 Length: 1434
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.09% Indels: 2
DB: 19 Gaps: 1

US-09-917-376-5 (1-88) x US-10-466-208-11 (1-1434)
QY 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 991 GGTATTCTGTACAATACAGACGGGATGGAGTATGAACAGCAACCAATCCGTCG 1050

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1051 CAGCTTCAATAAATAAATAACGGCAATACCAAGTTGATTTAAAGATGTCACTGCCGT 1110

QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 53
Db 1111 TACTGGTATAACGCGCAAAACAAAGCCCAAAACGTT-----GACTGTGACTACGCGCAG 1164

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1165 CTGTGATCGGCAATGTGCATACATAAGTTTGTGACGTTCATATAACCAAGCAAGTGCA 1224

QY 84 AspThrTyrLeuGln 88
Db 1225 GATACCTATCTGGAA 1239

RESULT 10
US-10-466-208-7
; Sequence 7, Application US/10466208
; Publication No. US20040180348A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 7
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCase gene with mutated signal sequence to enhance its
```

```
; OTHER INFORMATION: hydrophobicity
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1488)
US-10-466-208-7

Alignment Scores:
Pred. No.: 3,62e-16 Length: 1491
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.09% Indels: 2
DB: 19 Gaps: 1

US-09-917-376-5 (1-88) x US-10-466-208-7 (1-1491)
QY 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 1048 GGTATTCTGTACAATACAGACGGGATGGAGTATGAACAGCAACCAATCCGTCG 1107

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1108 CAGCTTCAATAAATAAATAACGGCAATACCAAGTTGATTTAAAGATGTCACTGCCGT 1167

QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaLa 63
Db 1168 TACTGGTATAACGCGCAAAACAAAGCCCAAAACGTT-----GACTGTGACTACGCGCAG 1221

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1222 CTGTGATCGGCAATGTGCATACATAAGTTTGTGACGTTCATATAACCAAGCAAGTGCA 1281

QY 84 AspThrTyrLeuGln 88
Db 1282 GATACCTATCTGGAA 1296

RESULT 11
US-10-433-577-34
; Sequence 34, Application US/10433577
; Publication No. US20040171065A1
; GENERAL INFORMATION:
; APPLICANT: Genofocus Co., Ltd.
; TITLE OF INVENTION: Method for Expression of Proteins on Spore Surface
; FILE REFERENCE: PCT-Genofocus-1/10/433,577
; CURRENT APPLICATION NUMBER: US/10/433,577
; CURRENT FILING DATE: 2003-06-05
; PRIOR APPLICATION NUMBER: KR2000-74835
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 34
; LENGTH: 2510
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (461)..(2491)
US-10-433-577-34

Alignment Scores:
Pred. No.: 7,25e-16 Length: 2510
Score: 188.00 Matches: 34
Percent Similarity: 61.18% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 31
Query Match: 40.09% Indels: 2
DB: 19 Gaps: 1

US-09-917-376-5 (1-88) x US-10-433-577-34 (1-2510)
QY 4 GlyVallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2051 GGTATTCTGTACAATACAGACGGGATGGAGTATGAACAGCAACCAATCCGTCG 2110
```



```
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 2111 CAGCTTCAATATAAATAACGCAATACCAACGCTTGATTTAAAAAGATGTCACCTCCCGT 2170
Qy 44 TyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrpAlaLa 63
Db 2171 TACTGGTATACGCGAATAAACAAGGCGCAAAACGTT-----GACTGTGACTACGCGAG 2224
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2225 CTTGATGCGCAATGTGCATACACAGTTTGTGAGTTGTGCATTAACCAACCAAGGTGCA 2284
Qy 84 AspThrTyrLeuGln 88
Db 2285 GATACCTATCTGGAA 2299

RESULT 12
US-10-933-404-1
; Sequence 1, Application US/10933404
; Publication No. US20050118729A1
; GENERAL INFORMATION:
; APPLICANT: MORAG, Ely
; TITLE OF INVENTION: MICROARRAYS OF CELLULOSE BINDING CHIMERIC PROTEINS AND METHODS OF
; FILE REFERENCE: THEREOF
; FILE REFERENCE: 85189-7100
; CURRENT APPLICATION NUMBER: US/10/933,404
; CURRENT FILING DATE: 2004-09-03
; PRIOR APPLICATION NUMBER: PCT/IL03/00177
; PRIOR FILING DATE: 2003-03-06
; PRIOR APPLICATION NUMBER: US 60/362,061
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-933-404-1

Alignment Scores:
Pred. No.: 1,62e-13 Length: 501
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 21 Gaps: 2

US-09-917-376-5 (1-88) x US-933-404-1 (1-501)
Qy 1 ValSerGlyValValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 13 GTATCAGGCAATTTGAAGTTGAAATCTACACAGCAATCTTCAGATACACTACTCA 72
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 73 ATCATCTCTCAGTTCAAGTTACTATATACCGAGCAAGTGCATTTGATTTGTCCAAATC 132
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 133 ACATTGAGATATATTATACAGTACAGCGACAGAAAGATCAGACCTTCTGG---TGTGAC 189
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleAlaAla 71
Db 190 CATGCTGCAATTAATCGGCGAGTAAACGCGAGTACACGGAATTTACTTCAATGTAAAGGA 249
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 250 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGCGACACCTACCTTGAA 300

RESULT 13
US-10-372-054-3
; Sequence 3, Application US/10372054
; Publication No. US20030203466A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. I633
US-10-372-054-3

Alignment Scores:
Pred. No.: 6,59e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 17 Gaps: 2

US-09-917-376-5 (1-88) x US-10-372-054-3 (1-1438)
Qy 1 ValSerGlyValValLeuValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTTGAAGTTGAAATCTACACAGCAATCTTCAGATACACTACTCA 999
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCTCTCAGTTCAAGTTACTATATACCGAGCAAGTGCATTTGATTTGTCCAAATC 1059
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATATTATATACAGTACAGCGACAGAAAGATCAGACCTTCTGG---TGTGAC 1116
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleAlaAla 71
Db 1117 CATGCTGCAATTAATCGGCGAGTAAACGCGAGTACACGGAATTTACTTCAATGTAAAGGA 1176
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGCGACACCTACCTTGAA 1227

RESULT 14
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
```



```
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; PRIOR APPLICATION NUMBER: 09/184,217
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-655-433-11

Alignment Scores:
Pred. No.: 6,86e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 18 Gaps: 2

US-09-917-376-5 (1-88) x US-10-655-433-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTTGAAGTTGAATCTACACAGCAATCTCTCAGATACTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTACTAATACCGAAGCAGTGCATTTGATTTGTCCAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATATACAGTAGACGAGCAAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TpdAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGGCAGCTACACGGAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGACAGACACCTACTCTTGAA 1278

Search completed: October 6, 2005, 05:10:28
Job time : 920.52 secs
```

```
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; PRIOR APPLICATION NUMBER: 09/184,217
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11

Alignment Scores:
Pred. No.: 6,86e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.79% Conservative: 23
Best Local Similarity: 36.08% Mismatches: 29
Query Match: 35.18% Indels: 10
DB: 13 Gaps: 2

US-09-917-376-5 (1-88) x US-10-072-152-11 (1-1482)
QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTTGAAGTTGAATCTACACAGCAATCTCTCAGATACTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTACTAATACCGAAGCAGTGCATTTGATTTGTCCAACTC 1110
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATATACAGTAGACGAGCAAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TpdAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGGCAGCTACACGGAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGACAGACACCTACTCTTGAA 1278

RESULT 15
US-10-655-433-11
; Sequence 11, Application US/10655433
; Publication No. US20040067572A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Samme O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20040067572A1el Pectate Lyases
; FILE REFERENCE: 5378,200-US
; CURRENT APPLICATION NUMBER: US/10/655,433
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
```